

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2002, 16:00:37 ; Search time 40.44 Seconds
(without alignments)
485.396 Million cell updates/sec

Title: US-09-925-122A-1

Perfect score: 1391

Sequence: 1 MKRTVSDNSLSNRGCKPD.....PTSYEVCLDKNAKAKTYI 265

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101:*

1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1391	100.0	265	20	AAV24922	Human SH3-containing protein sequ
2	1362.5	98.0	592	22	AAV94690	Human protein sequ
3	1345.5	96.7	537	22	AAV92528	Human protein sequ
4	782.5	56.3	367	22	AAE04187	Human gene 10 enco
5	777.5	55.9	547	22	AAE04186	Human NOVA2 protein
6	696.5	50.1	217	22	AAE04186	Human gene 10 enco
7	432.5	31.1	175	20	AAV24923	Human SH3-containing protein
8	274	19.7	211	20	AAV40582	Partial amino acid
9	274	19.7	211	20	AAV85101	Thyroid hormone re
10	274	19.7	212	15	AAV3543	Thyroid hormone re
11	274	19.7	212	20	AAV92389	Human TR-interacti

12	207	14.9	186	20	AAV60412	Human normal blad
13	196	14.1	249	22	AAV93133	Human protein sequ
14	155.5	11.2	441	17	AAV05410	Mouse H74 protein.
15	151.5	10.9	370	21	AAV99428	Human PRO1431 (UNQ
16	151.5	10.9	370	22	AAV66177	Protein of the inv
17	147.5	10.6	377	17	AAV05411	Human H74 protein.
18	147.5	10.6	486	21	AAV43338	Human OREF3102
19	147.5	10.6	486	21	AAV54041	Protein encoded by
20	145.5	10.5	506	21	AAV44242	Human cell signal1
21	141.5	10.2	424	21	AAV41838	Human OREF ORF1602
22	141.5	10.2	424	22	AAV27227	Human EXMAD-5 SBO
23	139.5	10.0	486	22	AAV24503	Human PAC amino ac
24	134.5	9.7	1214	21	AAV57444	Mouse Ees1 protein
25	134.5	9.7	1715	21	AAV57449	Mouse Ees1 protein
26	132	9.5	49	22	AAV35318	Peptide #935 enco
27	126	9.1	1197	21	AAV57445	Mouse Ees2 protein
28	126	9.1	1658	21	AAV57450	Mouse Ees2u protei
29	123.5	8.9	733	21	AAV39073	Human polypeptide
30	123.5	8.9	733	22	AAV40859	Human polypeptide
31	120	8.6	412	20	AAV49151	Amino acid sequenc
32	120	8.6	425	20	AAV49248	N-terminal region
33	120	8.6	425	20	AAV32187	Human polypeptide
34	118	8.5	1315	22	AAV22083	Human polypeptide
35	117	8.4	1269	22	AAV40297	Mouse SH3P13 prote
36	115.5	8.3	347	17	AAV05394	Rat phosphodiester
37	115.5	8.3	1683	17	AAV71160	Streptococcus pneu
38	115	8.3	453	19	AAV5081	Streptococcus pneu
39	115	8.3	641	20	AAV32158	Human SH3D1A prote
40	115	8.3	655	20	AAV92226	Cbpa of serotype 4
41	115	8.3	655	20	AAV32099	Choline binding pr
42	115	8.3	694	21	AAV81653	Streptococcus pneu
43	115	8.3	1035	22	AAV43519	Human polypeptide
44	115	8.3	1144	20	AAV32154	Human SH3D1A prote
45	115	8.3	1215	20	AAV32156	Human SH3D1A prote

ALIGNMENTS

RESULT 1	
AAV24922	AAV24922 standard; Protein: 265 AA.
XX	XX
AC	AAV24922;
XX	XX
DT	27-AUG-1999 (first entry)
XX	XX
DE	Human SH3-containing protein 1.
XX	XX
KW	Human SH3 containing protein; HS3C-1; HS3C-2; Src homology 3 domain;
KW	diagnosis: cancer; immune disorder; development disorder; leukemia;
KW	immunoflammatory condition; rheumatoid arthritis; ulcerative colitis;
KW	osteoarthritis; gaucher's disease; adenocarcinoma; lymphoma; melanoma;
KW	sarcoma; AIDS; allergy; asthma; irritable bowel syndrome; pancreatitis;
KW	multiple sclerosis; osteoarthritis; haemodialysis; infection; trauma;
KW	anaemia; epilepsy; congenital glaucoma.
XX	XX
OS	Homo sapiens.
XX	XX
PN	US916753-A.
XX	XX
PD	29-JUN-1999;
XX	XX
PF	13-NOV-1997; 97US-0970133.
XX	XX
PR	13-NOV-1997; 97US-0970133.
XX	XX
PA	(INCY-) INCYTE PHARM INC.
XX	XX
PI	Bandman O, Guegler KJ, Lal P;
XX	XX
DR	WPI, 1999-394206/33.
DR	N-PSDB; AAV3628.

XX New Src homology 3 domain containing proteins useful for the
PT diagnosis, treatment or prevention of cancer and immune or
XX development disorders
XX

PS Claim 1: Fig 1: 32pp: English.

XX The present sequence represents human Src homology 3 domain (SH3)
CC containing protein 1, designated HS3C-1. HS3C proteins can be used
CC for the diagnosis, treatment or prevention of cancer and immune or
CC development disorders. HS3C-1 is particularly expressed in prostate
CC tissues associated with prostate tumours and HS3C-2 with inflammatory
CC conditions such as rheumatoid arthritis, ulcerative colitis,
CC osteoarthritis and Gaucher's disease. A vector expressing the complement
CC of the polynucleotide encoding HS3C-1 can be administered to a subject
CC to prevent or treat cancers including adenocarcinoma, leukemia, lymphoma,
CC melanoma, sarcoma and especially cancers of the bladder, kidney, heart,
CC lung, adrenal gland, skin, spleen, liver, ovary, pancreas, thyroid and
CC uterus. An immune disorder such as AIDS, allergies, asthma, irritable
CC bowel syndrome, multiple sclerosis, pancreatitis and osteoarthritis can
CC also be treated along with complications of cancer, haemodialysis, viral,
CC bacterial, fungal, and parasitic infections and trauma. A vector
CC expressing the complement of the polynucleotide encoding HS3C-2 can also
CC be administered to a subject to prevent or treat cancers and immune
CC disorders as well as developmental disorders such as anaemia, epilepsy,
CC and congenital glaucoma. The expression vectors which encode HS3C can be
CC used to deliver nucleotide sequences to targeted organ, tissue or cell
CC populations and antisense polynucleotides to treat conditions associated
CC with overexpression of HS3C by blocking transcription of the mRNA,
CC modulating HS3C activity or regulating the gene function.
XX

SO Sequence 265 AA:

Query Match 100.0%; Score 1391; DB 20; Length 265;
Best Local Similarity 100.0%; Pred. No. 1.2e-115;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRTVS DNSLSNSRGEGRPDLKFGSKSGKGLWPFITKKNKGATPEDFS NLPDRKKKIQ 60
DB 1 MKRTVS DNSLSNSRGEGRPDLKFGSKSGKGLWPFITKKNKGATPEDFS NLPDRKKKIQ 60
QY 61 KYDELNKRIOEMODRDAITKMKDYVLTKNPOMGDPASLDHKLAEVSONIEKLRVETQKFE 120
DB 61 KYDELNKRIOEMODRDAITKMKDYVLTKNPOMGDPASLDHKLAEVSONIEKLRVETQKFE 120
QY 121 AMLAEVEGRLPARNRQARQSGLYDSQNPPTVNNCAQDRSPDGSYTEEOQSEEMKYLVA 180
DB 121 AMLAEVEGRLPARNRQARQSGLYDSQNPPTVNNCAQDRSPDGSYTEEOQSEEMKYLVA 180
QY 181 TDPDFDEPDEPLPAITGCTKALYTFEGONEGTISVBEGETLYVIEDKGDGWTIRRNED 240
DB 181 TDPDFDEPDEPLPAITGCTKALYTFEGONEGTISVBEGETLYVIEDKGDGWTIRRNED 240
QY 241 EEGYVPTS YVEVCLDKNNAKGAKTYI 265
DB 241 EEGYVPTS YVEVCLDKNNAKGAKTYI 265

RESULT 2

AAB94690 ID AAB94690 standard; Protein: 592 AA.

XX AAB94690:

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:15657.

XX Human: primer: detection; diagnosis; antisense therapy; gene therapy.

XX Homof. sapiens.

PN EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX

PS Claim 8: SEQ ID 15657; 2537bp + CD ROW; English.

XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences: AAB92446 to
XX AAB95893 represent human amino acid sequences: AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.

SO Sequence 592 AA:

Query Match 98.0%; Score 1362.5; DB 22; Length 592;
Best Local Similarity 87.7%; Pred. No. 1.3e-112;
Matches 265; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

QY 1 MKRTVS DNSLSNSRGEGRPDLKFGSKSGKGLWPFITKKNK----- 39

DB 291 MKRTVS DNSLSNSRGEGRPDLKFGSKSGKGLWPFITKKNKSPKQKPSHRTNEMTSKP 350

QY 40 -----GATPEDFS NLPDRKKKIQKYDELNKRIOEMODRDAITKMK 83

DB 351 KIHCFRS LKRGSLKLGATPEDFS NLPDRKKKIQKYDELNKRIOEMODRDAITKMK 410

QY 84 DYVLKNPOMGDPASLDHKLAEVSONIEKLRVETQKFEAMLAEEGRLPARNRQARQSG 143

DB 411 DYVLKNPOMGDPASLDHKLAEVSONIEKLRVETQKFEAMLAEEGRLPARNRQARQSG 470

QY 144 YDSQNPPTVNNCAQDRSPDGSYTEEOQSEEMKYLATPDFDEPDEPLPAITGCTKALY 203

DB 471 YDSQNPPTVNNCAQDRSPDGSYTEEOQSEEMKYLATPDFDEPDEPLPAITGCTKALY 530

QY 204 TFEONEGTISVBEGETLYVIEDKGDGWTIRRNED EGYVPTS YVEVCLDKNNAKGAKT 263

```

Db      531 tteggagctisvvegetlyvveedkgdgywtrirrmeeedgyvptsvvevcldnakgakt 590
Qy      264 YI 265
        ||
Db      591 yI 592

RESULT 3
AAB92528
ID      AAB92528 standard; Protein: 537 AA.
XX
AAB92528;
XX
26-JUN-2001 (first entry)
XX
Human protein sequence SEQ ID NO:10684.
XX
Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
Homo sapiens.
XX
EP1074617-A2.
XX
07-FEB-2001.
XX
28-JUL-2000; 2000EP-0116126.
XX
29-JUL-1999; 99JP-0248036.
XX
27-AUG-1999; 99JP-0300253.
XX
11-JAN-2000; 2000JP-0118776.
XX
02-MAY-2000; 2000JP-0183767.
XX
09-JUN-2000; 2000JP-0241899.
XX
PA
(HELI-) HELIX RES INST.
XX
Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
WPI: 2001-318749/34.
XX
Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
Claim 8: SEQ ID 10684; 2537bp + CD ROM; English.
XX
The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
Sequence 537 AA;
SQ

```

```

Query Match          96.7%: Score 1345.5; DB 22; Length 537;
Best Local Similarity 81.0%: Pred. No. 3.6e-111;
Matches 264; Conservative 1; Mismatches 0; Indels 61; Gaps 1;

Qy      1 MKRTVSDNSLSNRGEGKFDLKFSGKSGKGLMPPIKRNK----- 39
        |||
Db      212 mktvdsdnlslnsrggkfpdlkfsgkskglmpfllknklmslltspbqpppppasasp 271
        |||
Qy      40 -----GATPDDTSNLPPEGRKKLQ 59
        |||
Db      272 savpnpqspkqkqkplshrfnefmskpklnclfrslergatpedsnlppgqrklq 331
        |||
Qy      60 OKVDELNKEIQKEMDQRPATIKMKDVYLYKNPQMDPASLDHKLAEVSONIKLVEYRQKE 119
        |||
Db      332 qvdeImkeiqkemdqralkmkdvylknpmqdpasldhklavsgnklkivegkfk 391
        |||
Qy      120 EAWLAEEGRLPARNEQARQSGLYDSQNPPTVNNCAQDRSPGSYTEQSQSESKVL 179
        |||
Db      392 eawlaeevgrlparseqarrrqsglydsqnpptvnncaqdrspgsgyteeqsgsemkvl 451
        |||
Qy      180 ATDPDDEDFDDEEPLPATITCKALYTFEGQNGCTISVSGETLYIIEBKDGGMTRIRNE 239
        |||
Db      452 atdfderfdeepIpaIgtckalytfegqngctisvvegetlyvveedkgdgywtrirrne 511
        |||
Qy      240 DEEGYVPTSVYEVCLDNKNAKGAITYI 265
        |||
Db      512 deegyvpstsvyevcldnknakgaktYI 537
        |||

```

```

RESULT 4
AAE04187
ID      AAE04187 standard; Protein: 367 AA.
XX
AAE04187;
XX
09-AUG-2001 (first entry)
XX
Human gene 10 encoded secreted protein fragment, SEQ ID NO:179.
XX
Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
KW inflammation; neurological disorder; Alzheimer's disease; food additive;
KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
KW cell culture; chemotaxis; vulnerrary; binding partner identification;
KW gene therapy; chromosome 1.
XX
Homo sapiens.
XX
WO200134643-A1.
XX
17-MAY-2001.
XX
08-NOV-2000; 2000WO-US30629.
XX
12-NOV-1999; 990US-0164825.
XX
03-AUG-2000; 2000US-0222904.
XX
(HUMA-) HUMAN GENOME SCI INC.
XX
Ruben SM, Komatsoulis GA, Soppet DR, Shi Y;
PI WPI: 2001-374441/39.
XX
Nucleic acids encoding 24 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. Gaucher's disease,
PT Alzheimer's disease, Schmitzer syndrome, Creutzfeldt-Jacob disease,
XX diabetes mellitus and multiple sclerosis -
XX

```

PS Disclosure; Page 31-32; 532pp; English.

XX AAD08404-AAD08478 represent cDNAs corresponding to 24 human secreted
 CC protein genes, and AAE04100-AAE04170 represent the proteins they encode.
 CC AAE04172-AAE04197 represent human secreted protein fragments or variants.
 CC The secreted proteins and their genes are useful for preventing, treating
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.
 CC Pathological conditions can be diagnosed by determining the amount of the
 CC new protein in a sample or by determining the presence of mutations in
 CC the new genes. Specific uses are described for each of the 24 genes,
 CC based on the tissues in which they are most highly expressed, and include
 CC developing products for the diagnosis or treatment of proliferative
 CC disorders, cancer, tumours, foetal and developmental abnormalities,
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
 CC angiogenic disorders, kidney disorders, gastrointestinal disorders,
 CC pregnancy-related disorders, endocrine disorders, and infections. The
 CC proteins can also be used to aid wound healing and epithelial cell
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs
 CC before transplantation, for supporting cell culture of primary tissues,
 CC to regenerate tissues, to identify their cognate ligands or binding
 CC partners, and in chemotaxis, and can be used as a food additive or a
 CC preservative to modify storage properties. Antibodies specific for a
 CC protein of the invention can be used in alleviating symptoms associated
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The
 CC present sequence represents a human secreted protein fragment referred
 CC to in the disclosure of the invention.

XX Sequence 367 AA;

Query Match 56.3%; Score 782.5; DB 22; Length 367;
 Best Local Similarity 58.6%; Pred. No. 2.2e-61;
 Matches 156; Conservative 40; Mismatches 53; Indels 17; Gaps 6;

QY 3 RTVSDNSLSNRGE-GKPDLEFG-GSKSGKLMPTFKNNKGATPEDEFSNLPPQRRKKLQQ 60
 Db 112 rtisdgtlisaaskgesgkmdaktvpgkagkylwlgkpkpaledfshlppgrrkkllqg 171
 QY 61 KYDELNKELOKEMODRDAITTKKDYLYLNPKQMDPASLDHKLAEVSQNIETKRVETQKFE 120
 Db 172 rldelntrelqkesdqkdaInmkmdvyeKnpqmgdpslqpklaetmndrlrImehIhKne 231
 QY 121 AMLAEVEGRLLPARNQAROSGLYDSQNPPTVNN-CAQDRESPDSYTEEOSQE-----SE 175
 Db 232 awlsevegktigrqd--trhs-----dInhlvtqgrespgsytdanqevrppq 281
 QY 176 MKVLATDFDEDEEPLPAIGTCALYTFEGQNGTISVVEGETLYVIEEDKGDGWTRI 235
 Db 282 qhghnefdedddpripalghckaIypfdghnegtlamkegeVlylIeedkggwtr 341
 QY 236 RRNEDEEGYVPTSYVEVCLDKNAKGA 261
 Db 342 trqnggegyvptsyldvlelKnskys 367

RESULT 5
 AAB61130

ID AAB61130 standard; Protein: 547 AA.

AC AAB61130;

DT 30-MAR-2001 (first entry)

XX Human NOV2 protein.

KW Human; NOVX; antiinflammatory; cytosolic; neuroprotective;
 KM cerebroprotective; immunomodulator; vulnerrary; vasotrophic; gene therapy;
 KM hyperplasia; tumour; restenosis; psoriasis; Dupuytren's contracture;

KW diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease.

XX Homo sapiens.

OS WO200075321-A2.

PN 14-DEC-2000.

PD 01-JUN-2000; 2000WO-US15303.

PR 03-JUN-1999; 990S-0137322.

PR 16-MAR-2000; 2000US-0189810.

PR 22-MAR-2000; 2000US-0191158.

PR 30-MAR-2000; 2000US-0193086.

PR 31-MAY-2000; 2000US-0137322.

XX (CURA-) CURAGEN CORP.

XX Shinkets RA, Fernandes E, Herrman J, Vernet C;

DR WPI: 2001-102403/11.

DR N-PSDB: AAF27850.

XX New NOVX polypeptides and polynucleotides, useful in gene therapy, as a
 XX diagnostic marker, protein therapeutic, antibody or small molecule drug
 XX target for treating immune, proliferative and metabolic diseases and
 XX wound healing

PS Claim 1; Page 22-24; 194pp; English.

XX The present sequence is a new isolated polypeptide (NOVX). The NOVX
 CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for
 CC treating or preventing NOVX-associated disorders. They are also useful
 CC for determining the presence of or a predisposition to a disease
 CC associated with altered levels of the NOVX polypeptide or nucleic acid.
 CC These NOVX-associated disorders include hyperplasias, tumours,
 CC restenosis, psoriasis, Dupuytren's contracture, diabetic complications,
 CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral
 CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides
 CC are especially useful in gene therapy. Specifically, NOVX is useful as
 CC a diagnostic marker or prognostic marker, protein therapeutic and
 CC antibody target or small molecule drug target to treat disorders in the
 CC immune response pathway, thyroid and metabolic diseases, bone metabolic
 CC disorders, diseases of the pancreas (e.g. diabetes or digestive
 CC disorders), proliferative diseases, or tissue regeneration and
 CC development (e.g. wound healing or treatment of burns).

XX Sequence 547 AA;

Query Match 55.9%; Score 777.5; DB 22; Length 547;
 Best Local Similarity 58.3%; Pred. No. 1.1e-60;
 Matches 155; Conservative 41; Mismatches 53; Indels 17; Gaps 6;

QY 3 RTVSDNSLSNRGE-GKPDLEFG-GSKSGKLMPTFKNNKGATPEDEFSNLPPQRRKKLQQ 60
 Db 292 rtisdgtlisaaskgesgkmdaktvpgkagkylwlgkpkpaledfshlppgrrkkllqg 351
 QY 61 KYDELNKELOKEMODRDAITTKKDYLYLNPKQMDPASLDHKLAEVSQNIETKRVETQKFE 120
 Db 352 rldelntrelqkesdqkdaInmkmdvyeKnpqmgdpslqpklaetmndrlrImehIhKne 411
 QY 121 AMLAEVEGRLLPARNQAROSGLYDSQNPPTVNN-CAQDRESPDSYTEEOSQE-----SE 175
 Db 412 awlsevegktigrqd--trhs-----dInhlvtqgrespgsytdanqevrppq 461
 QY 176 MKVLATDFDEDEEPLPAIGTCALYTFEGQNGTISVVEGETLYVIEEDKGDGWTRI 235
 Db 462 qhghnefdedddpripalghckaIypfdghnegtlamkegeVlylIeedkggwtr 521
 QY 236 RRNEDEEGYVPTSYVEVCLDKNAKGA 261
 Db 522 trqnggegyvptsyldvlelKnskys 547

RESULT 6
 ID AAE04186 standard; Protein: 217 AA.
 AC AAE04186;
 DT 09-AUG-2001 (first entry)
 XX
 DE Human gene 10 encoded secreted protein fragment, SEQ ID NO:178.
 XX
 KW Human: secreted protein; proliferative disorder; cancer; tumour; asthma;
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 KW parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
 KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
 KW inflammation; neurological disorder; Alzheimer's disease; food additive;
 KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
 KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
 KW cell culture; chemotaxis; vulnerability; binding partner identification;
 KW gene therapy; chromosome 1.
 XX
 OS Homo sapiens.
 XX
 PN MO200134643-A1.
 XX
 PD 17-MAY-2001.
 XX
 PF 08-NOV-2000; 2000WO-US30629.
 XX
 PR 12-NOV-1999; 99US-0164825.
 XX
 PR 03-AUG-2000; 2000US-0222904.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Komatsoulis GA, Soppet DR, Shi Y;
 XX
 DR WPI: 2001-374441/39.
 XX
 PT Nucleic acids encoding 24 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating e.g. Gaucher's disease,
 PT Alzheimer's disease, Scimitar syndrome, Creutzfeldt-Jacob disease,
 PT diabetes mellitus and multiple sclerosis -
 XX
 PS Disclosure: Page 32; 532pp; English.
 XX
 CC AAD08404-AAD08478 represent cDNAs corresponding to 24 human secreted
 CC protein genes, and AAE04100-AAE004170 represent the proteins they encode.
 CC AAE04172-AAE04197 represent human secreted protein fragments or variants.
 CC The secreted proteins and their genes are useful for preventing, treating
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.
 CC Pathological conditions can be diagnosed by determining the amount of the
 CC new protein in a sample or by determining the presence of mutations in
 CC the new genes. Specific uses are described for each of the 24 genes,
 CC based on the tissues in which they are most highly expressed, and include
 CC developing products for the diagnosis or treatment of proliferative
 CC disorders, cancer, tumours, foetal and developmental abnormalities,
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
 CC angiogenic disorders, kidney disorders, gastrointestinal disorders,
 CC pregnancy-related disorders, endocrine disorders, and infections. The
 CC proteins can also be used to aid wound healing and epithelial cell
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs
 CC before transplantation, for supporting cell culture of primary tissues,
 CC to regenerate tissues, to identify their cognate ligands or binding
 CC partners, and in chemotaxis, and can be used as a food additive or
 CC preservative to modify storage properties. Antibodies specific for a
 CC protein of the invention can be used in alleviating symptoms associated
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,

CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The
 CC present sequence represents a human secreted protein fragment referred
 CC to in the disclosure of the invention.

XX
 XX
 SQ Sequence 217 AA;

Query Match 50.1%; Score 696.5; DB 22; Length 217;
 Best Local Similarity 59.0%; Pred. No. 4,7e-54;
 Matches 134; Conservative 35; Mismatches 43; Indels 15; Gaps 4;

OY 40 GATPEDFSNLPPEQRKKLQQAQVDEINKEIQEMDQDAITKKRDVYLKKNQKGPASLD 99
 1 gpladftshppqrkkllqgridelnrelqkesdqkdaInkdkyexkngpmslq 60
 DB 100 HKIAEVSQNIETKRVETQFEALAEVGLPARNEDQARQSLYSQNPPTVNN-CAQD 158
 61 pklaelmnmldrlmelhknawlsevgktggrgd--trhs2-----dinhlvttg 110
 OY 159 RESPDGTYTEGSOE---SEMKVLATDPDEFDPDEPPLPAIGTCALYTFEGONECTIS 214
 111 respegstydangevrrppqghnnefdedeodpIpaighckaiyfgdhnegtlia 170
 OY 215 VVEGETLYVIEEDKGDGQWTRIRRNDEDEGYVPTSYEVCIDKNAKGA 261
 171 mkegevllyieedkgdgtwrrarnggegyvptsyidvtleknskgs 217
 DB

RESULT 7
 ID AAY24923 standard; Protein: 175 AA.
 AC AAY24923;
 DT 27-AUG-1999 (first entry)
 XX
 DE Human SH3-containing protein 2.
 XX
 KW Human SH3 containing protein: HS3C-1; HS3C-2; Src homology 3 domain;
 KW diagnosis; cancer; immune disorder; development disorder; leukemia;
 KW immunoflammatory condition; rheumatoid arthritis; ulcerative colitis;
 KW osteoarthritis; Gaucher's disease; adenocarcinoma; lymphoma; melanoma;
 KW sarcoma; AIDS; allergy; asthma; irritable bowel syndrome; pancreatitis;
 KW multiple sclerosis; osteoarthritis; haemodialysis; infection; trauma;
 KW anaemia; epilepsy; congenital glaucoma.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 65 /label= unknown
 FT /note= "encoded by GAN"
 XX
 PN US5916753-A.
 XX
 PD 29-JUN-1999.
 XX
 PF 13-NOV-1997; 97US-0970133.
 XX
 PR 13-NOV-1997; 97US-0970133.
 XX
 PA (INCY-) INCYTE PHARM INC.
 XX
 PI Bandman O, Guegler KJ, Lal P;
 XX
 DR WPI: 1999-394206/33.
 DR N-PSDB; AAX83629.
 XX
 PT New Src homology 3 domain containing proteins useful for the
 PT diagnosis, treatment or prevention of cancer and immune or
 PT development disorders
 XX
 PS Claim 8; Fig 2; 32pp; English.

xx The present sequence represents human Src homology 3 domain (SH3)
cc containing protein 2, designated HS3C-2. HS3C proteins can be used
cc for the diagnosis, treatment or prevention of cancer and immune or
cc development disorders. HS3C-1 is particularly expressed in prostate
cc tissues associated with prostate tumours and HS3C-2 with immunoinflammatory
cc conditions such as rheumatoid arthritis, ulcerative colitis,
cc osteoarthritis and Gaucher's disease. A vector expressing the complement
cc of the polynucleotide encoding HS3C-1 can be administered to a subject
cc to prevent or treat cancers including adenocarcinoma, leukemia, lymphoma,
cc melanoma, sarcoma and especially cancers of the bladder, kidney, heart,
cc lung, adrenal gland, skin, spleen, liver, ovary, pancreas, thyroid and
cc uterus. An immune disorder such as AIDS, allergies, asthma, irritable
cc bowel syndrome, multiple sclerosis, pancreaticitis and osteoarthritis can
cc also be treated along with complications of cancer, haemodialysis, viral,
cc bacterial, fungal, and parasitic infections and trauma. A vector
cc expressing the complement of the polynucleotide encoding HS3C-2 can also
cc be administered to a subject to prevent or treat cancers and immune
cc disorders as well as developmental disorders such as anaemia, epilepsy,
cc and congenital glaucoma. The expression vectors which encode HS3C can be
cc used to deliver nucleotide sequences to targeted organ, tissue or cell
cc populations and antisense polynucleotides to treat conditions associated
cc with overexpression of HS3C by blocking transcription of the mRNA,
cc modulating HS3C activity or regulating the gene function.

SQ Sequence 175 AA;

Query Match	31.1%	Score 432.5	DB 20	Length 175
Best Local Similarity	49.5%	Pred. No. 9.8e-31		
Matches 91, Conservative	28	Mismatches 46	Indels 19	Gaps 4

QY 82 MKVDYLNKPKQGDPSADIDHKIAEVSQNIETKRYTQCFEAMIAEVEGRILPARNQARQ8 141

Db 1 mkvdyeekppqmgdpaslepqiaetlsmierkllevqkyeaawlacaesrv----lsmrgd 55

QY 142 GLYDSQNP-----TVNNCAQORESPDGSYTEEOSOSEMKVIATDFDFDEPDEE 191

Db 56 slsrharppxppasapddssnsnsdqtk-----sseeppceesgdtriyelfedf---eee 111

QY 192 PLPAIGTCKALTYEGQNEGTISVVEGETIYVIEEDKGDGDMTRIRNDEDEGVYPTSYVE 251

Db 112 ptpslgncvalyhtegsegstlmaegedslmeedkgdgdwtvrrtkeggegyvpslyr 171

QY 252 VCLD 255

Db 172 vcln 175

RESULT 8

ID AAY40582 standard; protein; 211 AA.

AC AAY40582;

DT 03-DEC-1999 (first entry)

Partial amino acid sequence of TR-interacting protein S410a

KW JLI protein; thyroid hormone receptor; trap assay; therapeutic;
KW nuclear hormone receptor protein; TR-interacting protein; human;
KW thyroid disorder.

OS	Homo sapiens.	
XX		
PN	US5962256-A.	
XX		
PD	05-OCT-1999.	
XX		
PF	06-JUN-1995;	95US-0471613
XX		
PR	04-Apr-1994;	94US-0222719
PR	30-OCT-1992;	92US-0596136

XX
PA (GEHO) GEN HOSPITAL CORP.
NY

PI Lee JW, Moore DD,

DR WPI; 1999-571268/48.

DR N-PSDB; AAZ07568.

PT Purified DNA comprising a sequence encoding a protein which
PT specifically interacts with a thyroid hormone receptor -
xx

PS Disclosure; Fig 14; 68pp; English.

CC The invention provides a JLI protein which specifically interacts with
CC thyroid hormone receptor in an *in vivo* trap assay. The JLI protein can
CC be recombinantly produced by standard recombinant methodology. The JLI
CC protein is used in a trap assay for determining whether a test protein
CC is capable of interacting with a nuclear hormone receptor protein, and
CC may also be used as a therapeutic peptide for treating thyroid
CC disorders. Sequences AY40577-556 represent partial amino acid sequences
CC of thyroid hormone receptor (TR)-interacting proteins.

Sequence 211 AA;

Query Match	19.7%;	Score 274;	DB 20;	Length 211;
Best Local Similarity	53.4%;	Pred. No. 1.5e-16;		
Matches 55; Conservative	18;	Mismatches 26;	Indels 4;	Gaps 2

```
OY      153   NNCADRRSPGSGSYEEESOSEKKVLTATDDEDDPEALPGCTKALTPEFGNEGT    212
       : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      17   nsaaqdtc-----sseepseesgdpilyrteldaf-eeepsighncvalyhfessegT    72

OY      213   ISVVGELTYIIEEDKGAGDMIRIRNDEDGGTVPTSTVENCLD    255
       ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db      73   lsmaegedslsimeekkgdgwlrvtvrickgeggvyplstlyrvln    115
```

RESULT 9

ID	AAW85101 standard; Protein; 211 AA
1	MA
2	MA
3	MA
4	MA
5	MA
6	MA
7	MA
8	MA
9	MA
10	MA
11	MA
12	MA
13	MA
14	MA
15	MA
16	MA
17	MA
18	MA
19	MA
20	MA
21	MA
22	MA
23	MA
24	MA
25	MA
26	MA
27	MA
28	MA
29	MA
30	MA
31	MA
32	MA
33	MA
34	MA
35	MA
36	MA
37	MA
38	MA
39	MA
40	MA
41	MA
42	MA
43	MA
44	MA
45	MA
46	MA
47	MA
48	MA
49	MA
50	MA
51	MA
52	MA
53	MA
54	MA
55	MA
56	MA
57	MA
58	MA
59	MA
60	MA
61	MA
62	MA
63	MA
64	MA
65	MA
66	MA
67	MA
68	MA
69	MA
70	MA
71	MA
72	MA
73	MA
74	MA
75	MA
76	MA
77	MA
78	MA
79	MA
80	MA
81	MA
82	MA
83	MA
84	MA
85	MA
86	MA
87	MA
88	MA
89	MA
90	MA
91	MA
92	MA
93	MA
94	MA
95	MA
96	MA
97	MA
98	MA
99	MA
100	MA
101	MA
102	MA
103	MA
104	MA
105	MA
106	MA
107	MA
108	MA
109	MA
110	MA
111	MA
112	MA
113	MA
114	MA
115	MA
116	MA
117	MA
118	MA
119	MA
120	MA
121	MA
122	MA
123	MA
124	MA
125	MA
126	MA
127	MA
128	MA
129	MA
130	MA
131	MA
132	MA
133	MA
134	MA
135	MA
136	MA
137	MA
138	MA
139	MA
140	MA
141	MA
142	MA
143	MA
144	MA
145	MA
146	MA
147	MA
148	MA
149	MA
150	MA
151	MA
152	MA
153	MA
154	MA
155	MA
156	MA
157	MA
158	MA
159	MA
160	MA
161	MA
162	MA
163	MA
164	MA
165	MA
166	MA
167	MA
168	MA
169	MA
170	MA
171	MA
172	MA
173	MA
174	MA
175	MA
176	MA
177	MA
178	MA
179	MA
180	MA
181	

AC AAW85101;

DT 09-FEB-1999 (first entry)

Thyroid hormone receptor-interacting protein S410a partial sequence.

KW Thyroid hormone receptor-interacting protein; S410a;
KW nuclear hormone receptor protein; screen.
WY

05 Homo sapiens

PN US5846711-A.
xy

PD 08-DEC-1998.

PF 04-APR-1994; 94US-0222719.

PR	04-APR-1994;	94US-0222719.
PR	30-OCT-1992;	92US-0969136.

PA (GEHO) GEN HOSPITAL, CORP.

PI. Lee JW, Moore DD;

DR WPI; 1999-059040/05.
N-DEDDA & AY93501

DR N-PSDB; AAV82591.

PT Screening assay for nuclear hormone receptor modulators - using
PT cells containing reporter gene construct

PS Disclosure; Fig 14; 69pp; English.

Best Local Similarity 53.4%; Pred. No. 1.5e-16;
Matches 55; Conservative 18; Mismatches 26; Indels 4; Gaps 2;

OY 153 NNCADRESDGVTYEQSESEKVLATDDEDEDEPLPATGTCALYTERGQNEGT 212

DB 17 nsasqdtke--ssepssesqdtplrtedelf-eeepslghcvaalyhtegsseg 72

OY 213 ISVVEGETLYVIEDKDGWTRIRNDEEGGYPTSYVEVCLD 255

DB 73 lsmaegdlslmeedkydgytwrvrkgegygypstlyrvltln 115

RESULT 12

AAV60412 standard; Protein; 186 AA.

AC AAV60412;

DT 31-JAN-2000 (first entry)

DE Human normal bladder tissue EST encoded protein 84.

KM Human; bladder; treatment; EST; expressed sequence tag; cytostatic;

OS Homo sapiens.

DE19818620-A1.

PD 28-OCT-1999.

PF 21-APR-1998; 98DE-1018620.

PR 21-APR-1998; 98DE-1018620.

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

DR WPI; 1999-602416/52.

DR N-PSDB; AA242176.

PT New polypeptides and their nucleic acids, useful for treatment of

PS Claim 23; Page 280; 366pp; German.

This invention describes novel polypeptide fragment sequences (I) and their encoding nucleic acids (II) which are highly expressed in normal bladder tissue and have cytostatic activity. (II) are used for recombinant expression of (I) and to isolate complete genes. (I) are used to identify agents suitable for the treatment of bladder tumours, to directly treat this form of cancer (including expression from gene therapy vectors), or are used in a preparation for cancer treatment. (I) is also used for the generation of specific antibodies. (II) are identified by assembling ESTs (expressed sequence tags) from a particular tissue type before comparison of expression patterns. This allows a significantly longer fragment of the gene to be revealed, and therefore reduces the number of failures because of ESTs from different libraries representing different parts of the same unknown gene. CC AA60329-160391 represent protein fragments encoded by the human normal bladder tissue cDNA library derived EST fragments represented in AA42122-242248.

CC Sequence 186 AA;

Query Match 14.9%; Score 207; DB 20; Length 186;

Best Local Similarity 100.0%; Pred. No. 1.1e-10;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MKRTVSQNSLSNSRGEGRDLKFGKSGKGLWPIFKKK 39

DB 136 mkrvtvsqnslnsrgegrdlkfgksgkglwpiFKKK 174

RESULT 13

AAAB93133 standard; Protein; 249 AA.

AC AAB93133;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:12027.

KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

PS Claim 8; SEQ ID 12027; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

CC Sequence 249 AA;

Query Match 14.1%; Score 196; DB 22; Length 249;

Best Local Similarity 100.0%; Pred. No. 1.6e-09;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRTVSDNLSNRGEGKPDLPKFGSKSGKLPFIKK 37
 |||
 Db 212 mkrtvdsnsisnrggkpkldkfgykskglwplfikk 248

RESULT 14
 AAM05410
 ID AAM05410 standard; Protein; 441 AA.
 AC AAM05410;
 XX
 DT 23-FEB-1998 (first entry)
 DE Mouse H74 protein.
 XX
 DE Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;
 KM cellular signaling element; cellular structural element; malignancy;
 KM protein identification; functional domain; protein screening;
 KW cellular signal transduction process.
 OS Mus musculus.
 XX

Key Location/Qualifiers
 FT Misc-difference 439
 FT /note="encoded by GAC"
 XX
 PN WO9631625-A1.
 XX
 PD 10-OCT-1996.
 XX
 PF 04-APR-1996; 96WO-US04454.
 XX
 PR 03-APR-1996; 96US-0630915.
 PR 07-APR-1995; 95US-0417872.
 XX
 PA (CYTO-) CYTOGEN CORP.
 PA (UYNC-) UNIV NORTH CAROLINA.
 XX
 PI Fowlkes DM, Hoffman N, Kay BK, Mcconnell SJ, Sparks AB;
 DR WPI; 1996-465045/46.
 DR N-PSDB; AAT39809.
 XX
 PT Identifying polypeptide(s) having specific functional domain (esp.
 PT SH3 domain) - comprises detecting selective binding to recognition
 PT unit, regardless of sequence homology
 XX
 PS Claim 102; Fig 47; 174pp; English.
 XX
 CC AAM05405-W05411 represent human and mouse Src-homology region 3 (SH3)
 CC domain containing proteins that can be used in the method of the
 CC invention. SH3 domain containing proteins play a role in signaling and
 CC structural elements of cells. The method of the invention is for
 CC identifying polypeptides containing functional domains of interest
 CC (especially SH3 domains). The method comprises contacting a multivalent
 CC recognition unit (RU) complex with a number of peptides and identifying
 CC polypeptides having a selective binding affinity for the RU complex. The
 CC method is based on functional similarities and does not rely on sequence
 CC similarities. Prior methods only gave limited success for identifying
 CC proteins which contain an SH3 domain due to the minimal sequence
 CC homology among known SH3 proteins. It has been found that small peptide
 CC RUS in multivalent form have reduced specificity for a given functional
 CC domain compared to monomer RUS. Multivalent RU complexes are particularly
 CC suited to screening for polypeptides containing functional domains that
 CC are similar to, but not identical in sequence to, the original target
 CC functional domain. The new method enables proteins having a common
 CC function to be identified. Identification of novel SH3 proteins will be
 CC useful for a better understanding of cell growth, malignancy, signal
 CC transduction processes, etc. New candidate drugs can be identified, and
 CC their specificities (e.g. pharmacological activities) can be assessed
 CC using the method of the invention.
 XX

SQ Sequence 441 AA;
 Query Match 11.2%; Score 155.5; DB 17; Length 441;
 Best Local Similarity 22.3%; Pred. No. 1.4e-05;
 Matches 64; Conservative 42; Mismatches 88; Indels 93; Gaps 11;

QY 36 KKNKGATPEDEFSNLPDQRRKKLOOKYDELNKEIQEKMDQDAITKMKDYLNKPQ---- 91
 | : ||| :
 Db 175 ktegsvrpe-----gqkklvdkvdrcqdvqktgkkyekv--ledvgktprpyvmeg 223

QY 92 -----MGDPASLDHKLAEVSONI-----EKLR- 113
 Db 224 megvfegcqgfeekrlvflkevllidkrlhlnlaensymhyrelegairgdaqedrlrw 283

QY 114 -----VETQKFAMVLAIEVSGRLPARMEQARROSGLYDSQNPPIYVNNCAQ--DRES 161
 Db 284 frstsgpmpmnpwqfgeewpdphtlakkckgqkaegatlslatgavestsqagdrts 343

QY 162 PDGSYTEEQSQSESMKYLATDFDDE-----FDDEEPLPAIGT-CKALYT 204
 | :
 Db 344 v-sgydrgqiyatews-----ddesgnpfgneangnampfedd----akgvrvralyd 392

QY 205 FEGQNEGTISVGEETLYVIEEDKGDGWTIRIRNEDEGVPTSYVE 251
 :
 Db 393 ydgqgdelsrfkagdeltklgedegqwcrgrlsdgsqglqlypanyve 439

RESULT 15
 AAY99428
 ID AAY99428 standard; Protein; 370 AA.
 XX
 AC AAY99428;
 XX
 DT 08-AUG-2000 (first entry)
 XX
 DE Human PRO1431 (UNQ737) amino acid sequence SEQ ID NO:315.
 XX
 DE Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
 KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
 OS Homo sapiens.
 OS
 XX WO200012708-A2.
 PN
 XX
 PD 09-MAR-2000.
 XX
 PF 01-SEP-1999; 99WO-US20111.
 XX
 PR 01-SEP-1998; 98US-0098716.
 PR 01-SEP-1998; 98US-0098749.
 PR 01-SEP-1998; 98US-0098750.
 PR 02-SEP-1998; 98US-0098803.
 PR 02-SEP-1998; 98US-0098821.
 PR 02-SEP-1998; 98US-0098843.
 PR 09-SEP-1998; 98US-0099536.
 PR 09-SEP-1998; 98US-0099596.
 PR 09-SEP-1998; 98US-0099598.
 PR 09-SEP-1998; 98US-0099602.
 PR 09-SEP-1998; 98US-0099642.
 PR 10-SEP-1998; 98US-0099741.
 PR 10-SEP-1998; 98US-0099754.
 PR 10-SEP-1998; 98US-0099763.
 PR 10-SEP-1998; 98US-0099792.
 PR 10-SEP-1998; 98US-0099808.
 PR 10-SEP-1998; 98US-0099812.
 PR 10-SEP-1998; 98US-0099815.
 PR 10-SEP-1998; 98US-0099816.
 PR 15-SEP-1998; 98US-0100385.
 PR 15-SEP-1998; 98US-0100386.
 PR 15-SEP-1998; 98US-0100390.
 PR 16-SEP-1998; 98US-0100584.
 PR 16-SEP-1998; 98US-0100627.

[illegible]

PR	30-OCT-1998;	98US-0106464.
PR	03-NOV-1998;	98US-0106856.
PR	03-NOV-1998;	98US-0106902.
PR	03-NOV-1998;	98US-0106905.
PR	03-NOV-1998;	98US-0106919.
PR	03-NOV-1998;	98US-0106932.
PR	03-NOV-1998;	98US-0106934.
PR	10-NOV-1998;	98US-0107783.
PR	17-NOV-1998;	98US-0108775.
PR	17-NOV-1998;	98US-0108779.
PR	17-NOV-1998;	98US-0108787.
PR	17-NOV-1998;	98US-0108788.
PR	17-NOV-1998;	98US-0108801.
PR	17-NOV-1998;	98US-0108802.
PR	17-NOV-1998;	98US-0108806.
PR	17-NOV-1998;	98US-0108807.
PR	17-NOV-1998;	98US-0108867.
PR	17-NOV-1998;	98US-0108925.
PR	18-NOV-1998;	98US-0108848.
PR	18-NOV-1998;	98US-0108849.
PR	18-NOV-1998;	98US-0108850.
PR	18-NOV-1998;	98US-0108851.
PR	18-NOV-1998;	98US-0108852.
PR	18-NOV-1998;	98US-0108858.
PR	18-NOV-1998;	98US-0108904.

(GETH) GENENTECH INC.

Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;

WPI; 2000-237871/20.

N-PSDB; AAA37110.

New mammalian DNA sequences encoding transmembrane, receptor or

small molecule inhibitors of the relevant receptor/ligand interactions

Claim 12; Fig 178; 773pp; English.

AAA37022 to AAA37144 encode the new isolated human transmembrane,

transmembrane and receptor PRO proteins can be used for screening of

receptor/ligand interactions. The polypeptides and nucleotide sequences

pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent

polypeptides from the present invention.

Sequence 370 AA;

Query Match 10.9%; Score 151.5; DB 21; Length 370;

Matches 60; Conservative 32; Mismatches 64; Indels 67; Gaps 11;

51. PEÖRRKKLÖQVDELNKEIQEKEMDÖRDAITKMKDVYLLKNPÖMGDPASLDHKLAEVSÖNIE 110

26 pnvqlqrlegr-----rqaserea-----pslegrlqevresir 60

111 KLRVETOKFEAWLAEEGR-----LPARN-----EQARRQSGLYDSQN--PPTVNN 154

61 raqvsqvkgaarlal1qgagldverwlkpaantqadevegerl1searlsqrdispt--- 117

155 CAQDRESPDGSYTEEQSESEMKVLATDFDDEFDDEEPLP-AIGT---CKA--LYTFEG 207

118 -aedaelsd---feceetgel-----feepapqalatraipcpahvfryqa 161

208 QNEGTSVVEGETLYVIEEDKGDGWTRIRNEDEEGVPTSYV 250

162 gredeltitegewleiegdadewkarnqhgevgfperyl 204

Mon Apr 8 06:09:04 2002

us-09-925-122a-1.rag

Page 11

Search completed: April 7, 2002, 16:08:16
Job time: 459 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2002, 16:07:42; Search time 40.07 Seconds
(without alignments)
967.361 Million cell updates/sec

Title: US-09-925-122A-1

Perfect score: 1391
Sequence: 1 MKRTVSDNSLSNRSRGCKPDK.....PTSYVEYCDKNKAKAKTYI 265

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_17:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mhc:*
8: sp.mammal:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.podent:*
13: sp.virus:*
14: sp.invertebrate:*
15: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1362.5	98.0	592	4	Q9H8H8
2	1345.5	96.7	537	4	Q9H8H8
3	1312	94.3	674	4	Q9H8H8
4	1068	76.8	237	11	Q61053
5	782.5	56.3	330	4	Q9H8H8
6	766	55.1	434	4	Q9H8H8
7	662	47.6	545	4	Q9H8H8
8	638	45.9	547	11	Q9H8H8
9	638	45.9	547	11	Q9H8H8
10	388.5	27.9	525	5	Q9H8H8
11	330	23.7	534	5	Q9H8H8
12	241.5	17.4	783	5	Q9H8H8
13	240.5	17.3	785	5	Q9H8H8
14	198.5	14.3	774	5	Q9H8H8
15	172.5	12.4	338	11	Q9H8H8
16	169	12.1	445	11	Q9H8H8
17	164.5	11.8	486	11	Q9H8H8
18	164.5	11.8	486	11	Q9H8H8
19	161.5	11.6	448	13	Q13154

20	159.5	11.5	434	4	Q9P2G8
21	159.5	11.5	444	4	Q9B111
22	159	11.4	445	4	Q9H0D3
23	158	11.4	447	11	Q9Q119
24	155.5	11.2	441	11	Q61644
25	155.5	11.2	441	11	Q920W5
26	154.5	11.1	701	4	Q9NXX8
27	153.5	11.0	488	11	Q9QX20
28	151	10.9	603	4	Q9P777
29	151	10.9	684	4	Q9A868
30	148	10.6	424	11	Q9EQP9
31	148	10.6	424	11	Q9Q3B8
32	147.5	10.6	335	4	Q9Y4V2
33	147.5	10.6	477	13	Q9DDA9
34	141.5	10.2	424	4	Q9UKS6
35	141.5	10.2	424	4	Q9H331
36	139.5	10.0	486	4	Q9UNF0
37	134.5	9.7	1714	11	Q920R4
38	129	9.3	687	11	Q9QY53
39	129	9.3	691	11	Q9QZP0
40	126	9.1	1197	11	Q920R5
41	126	9.1	1658	11	Q920R6
42	125.5	9.0	214	11	Q9CZV7
43	125.5	9.0	278	11	Q35104
44	125.5	9.0	312	11	Q9JKT0
45	124	8.9	427	2	Q9R8D9

ALIGNMENTS

RESULT 1
ID Q9H8H8 PRELIMINARY: PRT: 592 AA.
AC Q9H8H8:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CDNA FLJ13619.1 F1S. CLONE PLACE1010926, WEAKLY SIMILAR TO HYPOTHETICAL
DE 72.2 KD PROTEIN C12C2.05C IN CHROMOSOME II.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoka S., Ishii S., Kawai Y.,
RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,
RA Masuno Y., Kanehori K.;
RT "NEO human cDNA sequencing project";
RL Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AK023681; BAB14638.1; --
DR InterPro: IPR001060; FCH.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00611; FCH; 1.
DR Pfam: PF00018; SH3; 1.
DR SMART: SM00055; FCH; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50002; SH3; 1.
DR PROSITE: PS50002; SH3; 1.
SO SEQUENCE 592 AA: 68910 MW: 0CAC9D464706B667 CRC64;

Query Match 98.0%; Score 1362.5; DB 4; Length 592;
Best Local Similarity 87.7%; Pred. No. 8.5e-87;
Matches 265; Conservative 0; Mismatches 0; Indels 37; Gaps 1;
1 MKRTVSDNSLSNRSRGCKPDKFKGSKGKLPFKIRK-----39
|||||
Db 291 MKRTVSDNSLSNRSRGCKPDKFKGSKGKLPFKIRKNSPKQKKEPLSHRENEFTSKP 350

```
QY 40 -----GATPEDEFSNLPPEQRKKLQOKVDELNKELOKEMDQDAITKKM 83
DB 351 KIHCFSLRGLSLKLGATPEDEFSNLPPEQRKKLQOKVDELNKELOKEMDQDAITKKM 410
QY 84 DYLKRPQMGDPASLDHKLAEVSQNIKLEKLVYQKFEAWLAIEVGRLLPARNEQARRQSL 143
DB 411 DYLKRPQMGDPASLDHKLAEVSQNIKLEKLVYQKFEAWLAIEVGRLLPARNEQARRQSL 470
QY 144 YDSQNPPTVNNCAQDRSPDGSSTEEQOSEEMKVLATPDEDFDEEPLPAIGTCKALY 203
DB 471 YDSQNPPTVNNCAQDRSPDGSSTEEQOSEEMKVLATPDEDFDEEPLPAIGTCKALY 530
QY 204 TREGQNEGTISYVEGTLVYIEEDKGDGWTIRIRNDEEGYPTSYVEVCLDKNAGAKT 263
DB 531 TREGQNEGTISYVEGTLVYIEEDKGDGWTIRIRNDEEGYPTSYVEVCLDKNAGAKT 590
QY 264 YI 265
DB 591 YI 592
```

```
RESULT 2
Q9NMD1 PRELIMINARY: PRT: 537 AA.
ID 09NMD1.
AC 09NMD1.
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE HYDROTHERMAL 61.6 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-EMBRYO:
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoliri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Yamanebe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y.,
RA Niinomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK000975; BAA91451.1; -.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PSS0002; SH3; 1.
SQ SEQUENCE 537 AA: 61560 MW: CE14592678DD1A65 CRC64;
```

Query Match 96.7%; Score 1345.5; DB 4; Length 537;
Best Local Similarity 81.0%; Pred. No. 1,le-85;
Matches 264; Conservative 1; Mismatches 0; Indels 61; Gaps 1;

```
QY 1 MKRTVSDNSLSNRGEGKPDILKFGGSKGKLMPTIKKNK----- 39
DB 212 MKRTVSDNSLSNRGEGKPDILKFGGSKGKLMPTIKKNKLSLITSPHQPPPPASASP 271
QY 40 -----GATPEDEFSNLPPEQRKKLQ 59
DB 272 SAYVNGQSPKQOKEPILSHRENEFTWTSKPKIHCFRSLERGGATPEDEFSNLPPEQRKKLQ 331
QY 60 QKVDLKELOKEMDQDAITKKMDVYLKPNQMGDPASLDHKLAEVSQNIKLEKLVYQKFE 119
DB 332 QKVDLKELOKEMDQDAITKKMDVYLKPNQMGDPASLDHKLAEVSQNIKLEKLVYQKFE 391
QY 120 EAWLAIEVGRLLPARNEQARRQSLYDSQNPPTVNNCAQDRSPDGSSTEEQOSEEMKVL 179
DB 392 EAWLAIEVGRLLPARNEQARRQSLYDSQNPPTVNNCAQDRSPDGSSTEEQOSEEMKVL 451
```

```
QY 180 ATDPDEDFDEEPLPAIGTCKALYTFEGQNEGTISYVEGTLVYIEEDKGDGWTIRRNE 239
DB 452 ATDPDEDFDEEPLPAIGTCKALYTFEGQNEGTISYVEGTLVYIEEDKGDGWTIRRNE 511
QY 240 DEEGYPTSYVEVCLDKNAGAKTYI 265
DB 512 DEEGYPTSYVEVCLDKNAGAKTYI 537
```

```
RESULT 3
O60301 PRELIMINARY: PRT: 674 AA.
ID 060301.
AC 060301.
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE KIAA0554 PROTEIN (FRAGMENT).
DE KIAA0554.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN:
RX MEDLINE=98290545; PubMed=9628581;
RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N.,
RA Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. IX.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:31-39(1998).
DR EMBL; AB011126; BAA25480.1; -.
DR InterPro; IPR001060; FCH.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00611; FCH; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PROSITE; PSS0002; SH3; 1.
DR SMART; SM00055; FCH; 1.
DR SMART; SM00326; SH3; 1.
FT NON_TER 1
SQ SEQUENCE 674 AA: 77447 MW: D0B53BE0B2B8DE2E CRC64;
```

Query Match 94.3%; Score 1312; DB 4; Length 674;
Best Local Similarity 79.4%; Pred. No. 3,le-83;
Matches 258; Conservative 1; Mismatches 0; Indels 66; Gaps 1;

```
QY 1 MKRTVSDNSLSNRGEGKPDILKFGGSKGKLMPTIKKNK----- 39
DB 348 MKRTVSDNSLSNRGEGKPDILKFGGSKGKLMPTIKKNKLSLITSPHQPPPPASASP 407
QY 40 -----GATPEDEFSNLPPEQR 54
DB 408 SAYVNGQSPKQOKEPILSHRENEFTWTSKPKIHCFRSLRGLSLKLGATPEDEFSNLPPEQR 467
QY 55 RKLLQOKVDELNKELOKEMDQDAITKKMDVYLKPNQMGDPASLDHKLAEVSQNIKLEKLV 114
DB 468 RKLLQOKVDELNKELOKEMDQDAITKKMDVYLKPNQMGDPASLDHKLAEVSQNIKLEKLV 527
QY 115 ETQKFEAWLAIEVGRLLPARNEQARRQSLYDSQNPPTVNNCAQDRSPDGSSTEEQOSE 174
DB 528 ETQKFEAWLAIEVGRLLPARNEQARRQSLYDSQNPPTVNNCAQDRSPDGSSTEEQOSE 587
QY 175 EMKVATPDEDFDEEPLPAIGTCKALYTFEGQNEGTISYVEGTLVYIEEDKGDGWTIR 234
DB 588 EMKVATPDEDFDEEPLPAIGTCKALYTFEGQNEGTISYVEGTLVYIEEDKGDGWTIR 647
QY 235 IRNDEEGYPTSYVEVCLDKNNAK 259
DB 648 IRNDEEGYPTSYVEVCLDKNNAK 672
```

RESULT	4			
061053				
ID	061053	PRELIMINARY;	PRT;	237 AA.
AC	Q61053;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DJ	01-NOV-1996 (TREMBLrel. 01, Last sequence update)			
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	FORMIN BINDING PROTEIN FBP 17 (FRAGMENT).			
GN	FBNP1.			
OS	Mus musculus (Mouse).			
OC	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FVB;			
RX	MEDLINE=96183189; PubMed=8605874;			
RA	Chan D.C., Bedford M.T., Leder P.;			
RT	"Formin binding proteins bear WWP/MW domains that bind proline-rich			
RL	peptides and functionally resemble SH3 domains.";			
DR	EMBO J. 15:1045-1054(1996).			
DR	EMBL: U40751; AAC52479.1, "			
DR	HSSP; P15054; IBKM.			
DR	MGD; MGI:109606; Fmbp1.			
DR	InterPro: IPR001452; SH3.			
DR	Pfam: PF00018; SH3; 1.			
DR	PRINTS; PRO0452; SH3DOMAIN.			
DR	SMART; SM00326; SH3; 1.			
DR	Prosite; PSS0002; SH3; 1.			
DR	NON_TER			
FT	1			
SO	SEQUENCE	237 AA;	21173 MW;	4E23794C8CC744D7 CRC64;

	Query Match	76.8%	Score 1068:	Dt 11:	Length 237;
	Bates Local Similarity	90.0%;	Pred. No.	7.7e-67:	
	Matches_206; Conservative	5;	Mismatches	16;	Indels 0;
	Gaps				0
Oy	37 KKKGATPEDFSMLPEQORRRKTLQQKVDELINKEITQEMORDAITSKMKDVIYNKPQMCDPA	96		: :	:
Db	9 KKGGVTPDEPFSMPPEQRKRKLQAKVDLNRLEIQETTDGRDAITMKMDIYLNPKMGDPA	68		: :	:
Oy	97 SLDHKLAEYSNIEIKLRVETOKFEAMLAVEBSRLPARNEOAROSGLXDYSNPPIYYNNCA	156		: :	:
Db	69 SLDOLTEVTQNIEKLRLEAQKFEEAMLVAEESRQLPARSQAARSQSGLXDGTHQTWTNCA	128		: :	:
Oy	157 ODRESPDSGYTSEGOSESMEKYLANDDEDPERDPEDPLPATIGCKALTYPEGONEGTISVV	216		: :	:
Db	129 QORESPDSGYTSEGOSESHEKHVALPDDEFDFDEEPFLPATIGTKALTYEGONEGTISVV	188		: :	:
Oy	217 EGCTLVYEIEDKGDMWTRIRRNDEBEGYPITYSVCECLDKNMKGATTYI	265		: :	:
Db	189 EGETLSVIEEDKGDMWTRLRRNDEBEGYPTTSVYELDDKNMKGATTYI	237		: :	:
RESULT	5				
ID	GNNGGI PRELMINARY;	PRT:	330 AA.		
AC	GNNGGI.				
DT	01-OCT-2000 (TREMBLrel. 15, Created)				
JR	01-OCT-2000 (TREMBLrel. 15, last sequence update)				
DR	01-JUN-2001 (TREMBLrel. 17, last annotation update)				
DE	CDNA FLJ20275 FIB, CLONE HEP02372.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
CC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
OX	NCBII_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,				
RA	Nakashima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Ohayashi M.,				
RA	Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isogai T., Sugano S.;				
RT	"NEO human cDNA sequencing project."				

RL Submitted (FEB-2009) to the EMBL/GenBank/DBJ databases
 DR EMBL: AK000283; BA091051.1; -
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00325; SH3; 1.
 DR PROSITE: PS00002; SH3; 1.
 SO SEQUENCE: 330 AA; 37427MW; 2F5260C9FD3C46AF CRC64;

[illegible]

RESULT	6		
09BR51		PRELIMINARY;	PRT: 434 AA.
ID	09BR51		
AC	09BR51:		
DT	01-JUN-2001 (TEMBUREL, 17, Created)		
DT	01-JUN-2001 (TEMBUREL, 17, Last sequence update)		
DT	01-JUN-2001 (TEMBUREL, 17, Last annotation update)		
DE	DJ1033H22.1 (K1AA0554 PROTEIN) (FRAGMENT).		
GN	DJ1033H22.1.		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Wallis J.;		
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AL109613; CAC3651.1; --		
FT	NON_TER 1		
SQ	SEQUENCE 434 AA: 49389 MW; 2B12549604960040 CRC64;		

```
Query Match Similarity   55.1%; Score 766; DB 4; Length 43;
Best Local Similarity   49.5%; Pred. No. 1.3e-45;
Matches    160; Conservative    40; Mismatches    53; Indels    70; Gaps    7

QY      3 RTVSDNSLSNKSQE-GKPLDKFG-GSKRKLPEFIKKNGATP-----         43
| | | : : | : | | | | | | | | : |
Db     122 RITSDGTIASQSOSKMADKTTVGAKGKLMWLFGRPKRPQSLPTISLTSPNQSO 181

QY      44 -----EDFSNLPRERRRKKLQOKVDELNR        67
| | | | | | | | | | | | | | | | | | | |

Db     182 FLTFSIEPVHYCMNEIKTKGPRTPSFRSLKRGGLADEDSHPLEPGRRKKLDQRIDELNR 241

QY      68 EIQEKDORDDATTKMKDYVLKLNPMQMDPASLDHRLAEVSNIETKLVLPORKEAWLAEEV 127
| - | - | - | - | - | - | | | | | | | | | | | | | | | | | | |
Db     242 EIQQESDQDALNKMKDYEVKNQMOMDPSSLOPKLAEFMNNINIDLRLMEIHKNNAMWSLYE 301

QY      128 GRLLPAREGARROSGILYSOINPETVANN-QAODRESPDGSTTEEQOSE----SEMKYLAVTD 182
```

```

Db 302 GKTGSGD--RRHSS-----DINHLYQGRSESGSYTDPAQCEVGRPPQGHNHNE 351
OY 183 FDEDEDDEEPLPAIGTCALYTFEGONEGTISVVEGETLYVIEEDKGDGQWTRIRNEDEE 242
DB 352 FDEDEDDEEPLPAIGTCALYTFEGONEGTISVVEGETLYVIEEDKGDGQWTRARONGEE 411
OY 243 GYPTSYVEVCLDKNAKGAITYI 265
DB 412 GYPTSYIDVTELEKNSKGAITYI 434

RESULT 7
ID 015184 PRELIMINARY; PRT; 545 AA.
AC 015184;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JUN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CDC42-INTERACTING PROTEIN 4.
GN CIP4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97362357; PubMed=9210375;
RA Aspernstrom P.;
RT "A Cdc42 target protein with homology to the non-kinase domain of FER
RT has a potential role in regulating the actin cytoskeleton.";
RL Curr. Biol. 7:479-487(1997).
DR EMBL; AJ000414; CAA04062.1; -.
DR InterPro; IPR001060; FCH.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00611; FCH; 1.
DR PROSITE; PS00002; SH3; 1.
DR SMART; SM00055; FCH; 1.
DR SMART; SM00326; SH3; 1.
SQ SEQUENCE 545 AA; 62591 MW; 9C9D72EA734BC6E2 CRC64;

Query Match 47.6%; Score 662; DB 4; Length 545;
Best Local Similarity 52.3%; Pred. No. 2.8e-38;
Matches 136; Conservative 49; Mismatches 65; Indels 10; Gaps 6;

OY 1 MKRTVSDNSLSNRGEGKPDLKFGGSKGKLMPEFTKKNGATPEDEFSNLPEQRRKKLQ 60
DB 291 MNRAPSDSSL-GTPSDGRPELRGPRSRKRPFGKKNKTYVTEDEFSHLPEQQRKKLQ 349
OY 61 KVDLELNKEIQEMQORDAITYMKKDYLLKNPQMGDPASLDHKLAEVSQNIIEKLYETQKE 120
DB 350 QLEERSRLQKEVQRELAKKKKDYETKTPQMGDPASLEPQIAETLSNIERKLEQVQYE 409
OY 121 AMLAEVEGR-LPARNQARROSLYD--SQNPP--TVNNCAQDRSPDGSYTEQSOSE 175
DB 410 AMLAEESRVLNRSQDLSLRHARPPDPASAPPDSSSNSASQDTKE---SSDEPPESESQ 466
OY 176 MKVLATDDEDEDEEPLPAIGTCALYTFEGONEGTISVVEGETLYVIEEDKGDGQWTR 235
DB 467 DTPYLTEDDEF--DEEPSPIGHCAVYHFGSSSEGTISMAEGEDLSIMEEDKGDGQWTR 525
OY 236 RNEDEEGYPTSYVEVCLD 255
DB 526 RKEGEGYPTSYLRYTLN 545

RESULT 8
ID 015184 PRELIMINARY; PRT; 547 AA.
AC 015184;
DT 01-MAY-1997 (TREMBLrel. 03, Created)

```

```

DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SALT-TOLERANT PROTEIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=KIDNEY;
RA Tsuji E.;
RT Submitted (Aug-1997) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=KIDNEY;
RX MEDLINE=97112415; PubMed=8954095;
RA Tsuji E., Tsuji Y., Misumi Y., Fujita A., Sasaguri M., Ideishi M.,
RA Arakawa K.;
RT "Molecular cloning of a novel rat salt-tolerant protein by functional
RT complementation in yeast.";
RL Biochem. Biophys. Res. Commun. 229:134-138(1996).
DR EMBL; AB06914; BA02191.1; -.
DR InterPro; IPR001060; FCH.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00611; FCH; 1.
DR PROSITE; PS00002; SH3; 1.
DR SMART; SM00055; FCH; 1.
DR SMART; SM00326; SH3; 1.
SQ SEQUENCE 547 AA; 62787 MW; 16CF72FDCAA6B3F CRC64;

```

```

Query Match 45.9%; Score 638; DB 11; Length 547;
Best Local Similarity 50.6%; Pred. No. 1.3e-36;
Matches 132; Conservative 47; Mismatches 72; Indels 10; Gaps 6;

OY 1 MKRTVSDNSLSNRGEGKPDLKFGGSKGKLMPEFTKKNGATPEDEFSNLPEQRRKKLQ 59
DB 291 MNRVPSDSSLGTP--DGRLELAASSRSRAKRPFGKKNKTYVTEDEFSHLPEQQRKKLQ 348
OY 60 QVDELNKEIQEMQORDAITYMKKDYLLKNPQMGDPASLDHKLAEVSQNIIEKLYETQKE 119
DB 349 QLEERNRELQKEEDQRELAKKKKDYETKTPQMGDPASLEPQIAETLSNIERLNEVQRY 408
OY 120 EMLAEVEGR-LPARNQARROSLYD--SQNPPYNNCAQDRSPDGSYTEQ--SQS 174
DB 409 EMLAEETRVLSNRQDLSLRHARPPDPPTTAPPDSSSNSNSGSDNNMESSEEPPESEG 468
OY 175 EMKVLATDDEDEEPLPAIGTCALYTFEGONEGTISVVEGETLYVIEEDKGDGQWTR 234
DB 469 QDTPYLTEDDEF--DEPASPIGHCAVYHFGSSSEGTISMAEGEDLSIMEEDKGDGQWTR 526
OY 235 IRRNEDEEGYPTSYVEVCLD 255
DB 527 VRRKQGEYPTSYLRYTLN 547

RESULT 9
ID 099110 PRELIMINARY; PRT; 547 AA.
AC 099110;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE SIMILAR TO THYROID HORMONE RECEPTOR INTERACTOR 10.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (Feb-2001) to the EMBL/Genbank/DBJ databases.

```


RA Parsons J., Percy C., Rikfen L., Roopra A., Saunders D., Showmkeen R.,
 RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterson R.,
 RA Watson A., Weinstock L., Wilkinson-Spoat J., Wohlman P.;
 RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.*;
 RC Nature 368:32-38(1994).
 CC -1 SIMILARITY: TO EF-HAND FAMILY.
 DR EMBL: Z81568; CAB04595.1; -;
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR001060; FCH.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00036; efhand; 1.
 DR Pfam: PF00611; FCH; 1.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00055; FCH; 1.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS50002; SH3; 1.
 KW Calcium-binding.
 SQ SEQUENCE 785 AA; 88576 MW; F0DC36C4EA728508 CRC64;

Query Match 17.3%; Score 240.5; DB 5; Length 785;
 Best Local Similarity 23.7%; Pred. No. 6.9e-09;
 Matches 75; Conservative 56; Mismatches 108; Indels 77; Gaps 9;
 QY 8 NSLSNRSGGKPKDLKFGK-----SKGKL--WPEIKKNGKA--TPEDFSN----- 48
 DB 472 DSRTNDSADGS-----GGKLLSSPSKRNRIKRFGLIKEKEDEPEASNNQQLMYTK 526
 QY 49 -LPEQRRKRLQOKVDLKNKEIOKENDORDATKMDVYLKNPQMDPASLDKRLAEVSQ 107
 DB 527 SKPAHRLSCLRSKIRDMEOALQIGRGITRLQAVYTNPHQGNPSACTEPLISYAK 586
 QY 108 NIEKLFVEQKPEAMLAIEVGRIPARNEQA-----RRSGLYDSQNPPTVNNCAQ 157
 DB 587 KLEKLMKDHNKLEFYAMLEMSVEEQERSFGGRDTPDTRSMGSGTNOSSKTTEDVL 646
 QY 158 DRESPDGYSTEEOSOE-----SEMKVLATDFD 184
 DB 647 SEAGAGSSADSSKNILQOLFTRPKRLISSPTKSSSTPTPLKRAELSSPKILLSSFS 706
 QY 185 DEF-----DDEEPLDPAICTKALYTFEGONEGTISVVEGETLYVIEEDKGDGWT-- 1 235
 DB 707 GAIRKSLSTPDSYKVTATVTFALFEFAKSSAETMSIEDQELLVLEHHDGDMTRKNC 766
 QY 236 RKNDEEGYVPTSYE 251
 DB 767 RKHNESGEVPTSYLQ 782

RESULT 14
 QYVSUB PRELIMINARY; PRT; 774 AA.
 AC QYVSUB;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CG4684 PROTEIN (FRAGMENT).
 GN CG4684.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Tracheata; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCBI_Taxid=7227;
 ON 111
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertan B.P., Bhattacharya D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies A.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mostreli A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleib J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Styks R., Tector C., Turner R., Venter L., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gbbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT *The genome sequence of Drosophila melanogaster.*;
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003553; AAP5013.1; -;
 DR HSSP: P29354; IGR1.
 DR FlyBase: FBgn0035966; CG4684.
 DR InterPro: IPR001060; FCH.
 DR InterPro: IPR00108; Neu_cyt_fact_2.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00018; SH3; 2.
 DR PRINTS: PR00499; P67PHOX.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR SMART: SM00055; FCH; 1.
 DR SMART: SM00326; SH3; 2.
 DR PROSITE: PS50002; SH3; 2.
 DR NON_TER 774
 SQ SEQUENCE 774 AA; 87266 MW; 69641BICE27FB0CC CRC64;

Query Match 14.3%; Score 198.5; DB 5; Length 774;
 Best Local Similarity 24.2%; Pred. No. 5.5e-06;
 Matches 69; Conservative 32; Mismatches 93; Indels 91; Gaps 8;
 QY 56 KKLQOKVDLKNKEIOKEMDORDATKMDVYLKNPQMDP-----ASLDKRLAEVSQ 110
 DB 335 KNLGARGVKNENARENAKKLALQOSLRD-----SGQGTDPNDPGLDKRIEERFQIR 390
 QY 111 KLRVETOKFEA-----WLAEVG----- 128
 DB 391 RSETEKTKAEACLOCLRDGGINVDENVOEAEENMGVQLTRSSASSISRTDASGGENPSS 450
 QY 129 -----RLPARNDGARQSGLYDSQNPPTVNNCAQDRESPDGYSTEEOSQ 172
 DB 451 DSFYDSKKEETQAAAGTKPKQEOQLSRDRTFSDEDEPEV-----RPSAAASAAAS 504
 QY 173 ESEMKVLATDFDD-----EFDE-----EPLDPAICTKALYTFEGONEGTISV 216
 DB 505 SSMAWASAGWDPTTEVNMWAGGEEDDKDEPIVPEPEALFKCALVSYAQNDELTLIV 564
 QY 217 EGETLYVIEEDKGDGWTIRIRNDEEGYVPTSYEVLCKNAKA 261
 DB 565 ENOLEVYVGEAGDGLRARNRYRGEEGYVPHNYLDI--DOETAGS 607

RESULT 15
 Q99L37 PRELIMINARY; PRT; 338 AA.
 ID Q99L37
 AC Q99L37
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE SIMILAR TO KIA0554 PROTEIN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=MAMMARY TUMOR;
 RA Strausberg R;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC003867; AAH03867.1; -
 SQ SEQUENCE 338 AA; 39972 MW; F83CA4FE103657D2 CRC64;

Query Match 12.4%; Score 172.5; DB 11; Length 338;
 Best Local Similarity 84.6%; Pred. No. 0.00014;
 Matches 33; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

Qy 1 MKRTVSDNSLSNRGEGKPPDLKFGGSKGKGLWPFIRKKNK 39
 |||||
 Db 291 MKRTVSDNSLSNRGEGKPPDLKFGGSKGKGLWPFIRKKNK 328

Search completed: April 7, 2002, 16:13:11
 Job time: 329 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2002, 16:05:52 ; Search time 19.76 Seconds
(without alignments)
301.790 Million cell updates/sec

Title: US-09-925-122A-1

Perfect score: 1391

Sequence: 1 MKRTVSDNSLSNSGCKPD.....PTSYVEYLDKNAKATYI 265

Scoring table:

BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfillsl.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1391	100.0	265	2	US-08-970-133-1
2	1068	76.8	237	2	US-08-970-133-5
3	432.5	31.1	175	2	US-08-970-133-3
4	115	8.3	453	4	US-08-961-083-38
5	109.5	7.9	486	2	US-08-942-423-2
6	108.5	7.8	659	1	US-08-391-615-2
7	108.5	7.8	1151	3	US-08-840-006-6
8	108.5	7.8	1200	3	US-08-840-006-6
9	106	7.6	401	2	US-08-549-004-5
10	106	7.6	401	4	US-09-051-982A-5
11	105.5	7.6	659	5	PCT-US95-05008-8
12	103.5	7.4	554	4	US-09-308-022-6
13	102.5	7.4	553	1	US-08-473-894-2
14	102.5	7.4	553	1	US-08-484-710-2
15	102.5	7.4	553	1	US-08-484-709-2
16	102.5	7.4	553	4	US-08-474-697-2
17	102.5	7.4	1162	2	US-08-728-123A-2
18	102	7.3	1164	4	US-08-923-992A-10
19	101.5	7.3	536	1	US-07-820-011A-4
20	101.5	7.3	536	5	PCT-US93-00445-4
21	101.5	7.3	536	5	PCT-US95-05008-13
22	100	7.2	486	2	US-08-942-423-3
23	99.5	7.2	367	3	US-09-141-047-10
24	99.5	7.2	688	3	US-09-141-047-8
25	99	7.1	249	1	US-08-466-603-2
26	99	7.1	249	1	US-08-314-503A-2
27	99	7.1	249	1	US-08-468-066-2

28	99	7.1	249	2	US-08-466-717-2	Sequence 2, Appl1
29	99	7.1	249	2	US-08-766-738-4	Sequence 4, Appl1
30	99	7.1	249	3	US-08-466-743-2	Sequence 2, Appl1
31	99	7.1	249	5	PCT-US95-12414-2	Sequence 2, Appl1
32	99	7.1	415	3	US-08-938-830-1	Sequence 1, Appl1
33	99	7.1	415	3	US-09-020-222-1	Sequence 1, Appl1
34	99	7.1	1098	4	US-08-923-992A-8	Sequence 8, Appl1
35	98	7.0	60	1	US-08-627-497-1	Sequence 1, Appl1
36	98	7.0	533	1	US-07-820-011A-2	Sequence 2, Appl1
37	98	7.0	533	5	PCT-US93-00445-2	Sequence 2, Appl1
38	97.5	7.0	400	3	US-08-938-830-29	Sequence 29, Appl1
39	97	7.0	1128	4	US-08-923-992A-6	Sequence 6, Appl1
40	97	7.0	1164	4	US-08-923-992A-2	Sequence 2, Appl1
41	96	6.9	536	5	PCT-US95-05008-12	Sequence 12, Appl1
42	95.5	6.9	442	1	US-08-391-615-6	Sequence 6, Appl1
43	95	6.8	537	5	PCT-US95-05008-11	Sequence 11, Appl1
44	94.5	6.8	784	4	US-09-371-913A-7	Sequence 7, Appl1
45	94	6.8	251	2	US-08-766-738-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-08-970-133-1
; Sequence 1, Application US/08970133
; Patent No. 5916753
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Guegler, Karl J.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: SH3-CONTAINING PROTEINS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/970.133
; FILING DATE: Filed Herewith
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0419 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 265 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAT10703
; CLONE: 865744
; US-08-970-133-1

Query Match 100.0%; Score 1391; DB 2; Length 265;
Best Local Similarity 100.0%; Pred. No. 1.8e-113;
Matches 265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	KKRYSDNSLSNSREKGPDLTKFGKSKGKTLMPKTKKNGATTPEDFSNLPPEDQRRKKLQ	60
Dp	1	MKRYSDNSLSNSRQEGPDLTKFGKSKGKTLMPKTKKNGATTPEDFSNLPPEDQRRKKLQ	60
Qy	61	KVDELNKEIQKEMDORDAITKMKDVYLLKNPOMGDPASLDHKLAEVSQNIETLRVETQKE	120
Dp	61	KVDELNKEIQKEMORDAITKMKDVYLLKNPOMGDPASLDHKLAEVSQNIETLRVETQKE	120
Qy	121	AMLAEBERLRAFNDOARQSGLDLDSQNPPIVNNACADRESPODSTYIEBOSQSEMKVLA	180
Dp	121	AMLAEBERLRAFNDOARQSGLDLDSQNPPIVNNACADRESPODSTYIEBOSQSEMKVLA	180
Qy	181	TDFDDEFDDEEPLPAIGTCKALATFEQNEGTISVSGETLYVIEEDKGGWTRIRNED	240
Dp	181	TDFDDEFDDEEPLPAIGTCKALATFEQNEGTISVSGETLYVIEEDKGGWTRIRNED	240
Qy	241	EEGYVPTSYVEVCLDKNAKGAITYI 265	
Dp	241	EEGYVPTSYVEVCLDKNAKGAITYI 265	

```

1      RESULT 2
2      US-08-970-133-5
3      Sequence 5, Application US/08970133
4      Patent No. 5916753
5      GENERAL INFORMATION:
6      APPLICANT: Bandman, Olga
7      APPLICANT: Guegler, Karl J.
8      APPLICANT: Lal, Preeti
9      TITLE OF INVENTION: SH3-CONTAINING PROTEINASES
10     NUMBER OF SEQUENCES: 5
11     CORRESPONDENCE ADDRESS:
12     ADDRESSSEE: Incyte Pharmaceuticals, Inc.
13     STREET: 3174 Porter Dr.
14     CITY: Palo Alto
15     STATE: CA
16     COUNTRY: USA
17     ZIP: 94304
18     COMPUTER READABLE FORM:
19     MEDIUM TYPE: Diskette
20     COMPUTER: IBM Compatible
21     OPERATING SYSTEM: DOS
22     SOFTWARE: FASTSEQ for Windows Version 2.0
23     CURRENT APPLICATION DATA:
24     APPLICATION NUMBER: US/08/970,133
25     FILING DATE: Filed Herewith
26     PRIOR APPLICATION DATA:
27     APPLICATION NUMBER:
28     FILING DATE:
29     ATTORNEY/AGENT INFORMATION:
30     NAME: Billings, Lucy J.
31     REGISTRATION NUMBER: 36,749
32     REFERENCE/DOCKET NUMBER: PF-0419 US
33     TELECOMMUNICATION INFORMATION:
34     TELEPHONE: 650-855-0555
35     TELEFAX: 650-845-4166
36     INFORMATION FOR SEQ ID NO: 5:
37     SEQUENCE CHARACTERISTICS:
38     LENGTH: 237 amino acids
39     TYPE: amino acid
40     STRANDEDNESS: single
41     TOPOLOGY: linear
42     IMMEDIATE SOURCE:
43     LIBRARY: Genbank
44     CLONE: 1255033
45     IS-08-970-133-5

```

Query Match	76.8%;	Score 1068;	DB 2;	Length 237;
Best Local Similarity	90.0%;	Pred. No. 1.8e-85;		
Matches 206;	Conservative	5;	Mismatches 18;	Indels 0;
				Gaps 0;

Qy	37	KNKGATPEDFNSULPPEQORRRKLQOKVBEJLNEIKQEMDQPAITIKMDVYLAKNQMDPA	96
Db	9	KRGVATPEDFNSFPEQORRRKLQOKVVDLNEIKQETDQRAITIKMDVYLAKNQMDPA	68
Qy	97	SLDHIAEVSQNIETKLRYETQFEAMLAEEVGRLPARNEQARQSGLYDSQNPPTVYNCA	156
Db	69	SLDQKLTVEVQNIETKLRLREAQFEAMLAEEVGRLPARNSEQARRSGLYDQTHQTVYNCA	128
Qy	157	QDRSPDGSYTEEEOQSSEMKVLTATDPDEFDEEPLPAIGTCALYTFEGQNEGTSIV	216
Db	129	QDRSPDGSYTEEEOQSSEHVLVLPDPDEFDEEPLPAIGTCALYTFEGQNEGTSIV	188
Qy	217	EGETIYVEEDKGDGWRIRIRNEDEEGVPRSYVECLDNNAKAKYI	265
Db	189	EGETIYVEEDKGDGWRIRIRNEDEEGVPRSYVECLDNNAKAKYI	237

```

1      RESULT      3
2      US-08-970-133-3
3      ; Sequence 3, Application US/08970133
4      ; Patent No. 5916753
5      ;
6      ; GENERAL INFORMATION:
7      ; APPLICANT: Bandman, Olga
8      ; APPLICANT: Guegler, Karl J.
9      ; APPLICANT: Lal, Preeti
10     ; TITLE OF INVENTION: SH3-CONTAINING PROTEIN(S)
11     ; NUMBER OF SEQUENCES: 5
12     ; CORRESPONDENCE ADDRESS:
13     ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
14     ; STREET: 3174 Porter Dr.
15     ; CITY: Palo Alto
16     ; STATE: CA
17     ; COUNTRY: USA
18     ; ZIP: 94304
19     ;
20     ; COMPUTER READABLE FORM:
21     ; MEDIUM TYPE: Diskette
22     ; COMPUTER: IBM Compatible
23     ; OPERATING SYSTEM: DOS
24     ; SOFTWARE: FastSeq for Windows Version 2.0
25     ;
26     ; CURRENT APPLICATION DATA:
27     ; APPLICATION NUMBER: US/08/970,133
28     ; FILING DATE: Filed Herewith
29     ; PRIOR APPLICATION DATA:
30     ; APPLICATION NUMBER:
31     ; FILING DATE:
32     ; ATTORNEY/AGENT INFORMATION:
33     ; NAME: Billings, Lucy J.
34     ; REGISTRATION NUMBER: 36,749
35     ; REFERENCE/DOCKET NUMBER: PF-0419 US
36     ; TELECOMMUNICATION INFORMATION:
37     ; TELEPHONE: 650-855-0555
38     ; TELEFAX: 650-845-4166
39     ;
40     ; INFORMATION FOR SEQ. ID NO: 3:
41     ; SEQUENCE CHARACTERISTICS:
42     ; LENGTH: 175 amino acids
43     ; TYPE: amino acid
44     ; STRANDEDNESS: single
45     ; TOPOLOGY: linear
46     ; IMMEDIATE SOURCE:
47     ; LIBRARY: PROSNOT20
48     ; CLONE: 1816529
49     ;
50     ; US-08-970-133-3

```

Query Match	31.1%;	Score 432.5;	DB 2;	Length 175;
Best Local Similarity	49.5%;	Pred. No. 2e-30;		
Matches	91;	Conservative	28;	Mismatches 46;
				Indels 19;
				Gaps 4;

```

0Y      82 MKDYVYLNKPOQGDASLDHKLAEVSQNIETKRVETQCFEAMLAVERLRPARNEQARRS 141
        | | | | | | | | | | : : | | | : | | | | | | : |
        | | | | | | | | | | : : | | | : | | | | | | : |
        1 MKDYVETKPOQGDASLEPOIAETLSNIETKLEVOXYEAMLAEBERV-----LSNRGD 55

0Y      142 GLYDSQNP-----IYNNAQDRESBDSGYTTEQSQSESBMKLAIFDDEPDEDEE 191

```

Db 56 SLRHHAPRPASAPPPSSNSASQOTKE---SSEPPSESGDTIYTFEDDF-EE 111
QY 192 PLPAIGTCKALLYFEGONEGTISVGEETLYVIEEDKGDGTRIRRNDEEGYPTSYVE 251
Db 112 PTSPIGHCAIYHFEESGEGTISMAEGEDLSLMEDEKGDGTRVRRREGGYPTSYLR 171
QY 252 VCLD 255
Db 172 VTLM 175

RESULT 4
US-08-961-083-38
; Sequence 38, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 453 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-38

Query Match 8.3%; Score 115; DB 4; Length 453;
Best Local Similarity 21.8%; Pred. No. 0.027;
Matches 57; Conservative 45; Mismatches 108; Indels 52; Gaps 9;

QY 1 MKRTYSDNSLSNRGECFDPDKFGSKSGKMLPFI-----KKNKGATPDPFS----- 47
Db 22 LKKITDDEEAEKEAKKRAADAKEGGKPRGKRGVPGELATPPDKKENAKSSDSSVGEET 81
QY 48 ----NLPEQRKKLQOQVDELNK--ETQKEMDQDAITTKMDVYLKFNQMGDPASLDHK 101
Db 82 LPSPLEKPEKVAEAEKVEEAKKKAEDQKEDRRNYPF---NTY-----KTLDEL 129
QY 102 LAEVSQNTLEKLRVETQKFEAMLAEVEGRLPARNEQARRSGSLYDSQNPPT--VNNCAQDR 159
Db 130 IAESE-----VEYKKALELYKEAKEPRNEEKVKQAKAEVESKKAATLEKIKTR 182
QY 160 ESPDSYTEEQSESEMKVLAITDFDDEDFDEPLPAIGTCKALYTFEGONEGTISVEGE 219

Db 183 KKAEEAEARKKAEDKVK-----EKPAEQOPAPAPAKERKAPAPKPDNPAEQPKAEKP 236
QY 220 TLYVIEEDKGDGWRIRRNED 241
Db 237 ADQQAEDYA-----RRSEEE 252

RESULT 5
US-08-942-423-2
; Sequence 2, Application US/08942423
; Patent No. 5891673
; GENERAL INFORMATION:
; APPLICANT: Hashimoto, Yasuhiro
; APPLICANT: Takemoto, Yoshihiro
; TITLE OF INVENTION: Lck Binding Protein
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Syntex (U.S.A.) Inc.
; STREET: 3401 Hillview Ave.
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,423
; FILING DATE: 01-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,715
; FILING DATE: 23-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Perles, Rohan
; REGISTRATION NUMBER: 35,752
; REFERENCE/DOCKET NUMBER: 28260
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 852-1698
; TELEFAX: (415) 496-3529
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 486 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: HSI
; US-08-942-423-2

Query Match 7.9%; Score 109.5; DB 2; Length 486;
Best Local Similarity 21.0%; Pred. No. 0.089;
Matches 56; Conservative 41; Mismatches 87; Indels 83; Gaps 13;

QY 46 FSNLPPEQRKKLQOQVDELNKETQKEMQDAITKMDVYLKPNQ-----GDPASLDH 100
Db 242 FESLAEKKRKRREKEKAQOMARQO-----ERKAVYKMSR-EVQQPSMPVEERAPRQLFK 296
QY 101 KLAEVSQNTLEKLRVETQKFEAMLAEVEGRLPARNEQARR-----OSGLY 144
Db 297 KISS-----EYW-PPAESHLPPESQVRSRREVPYSLPTROSPLONHLE 340
QY 145 DSQNPPT-----VNNCAQDR--SPDSYTEEQSESEMKVLA---TDFD 184
Db 341 DNEEPALPPTPEGLQVVEEYEAAPLEPEPEPDYEPETEPDYEDVGEIDQDED 400

[illegible]

```

1      RESULT      6
2      US-08-391-615-2
3      ; Sequence 2, Application US/08391615
4      ; Patent No. 5550054
5      GENERAL INFORMATION:
6      APPLICANT:  Witte, Owen
7      APPLICANT:  Tsukada, Satoshi
8      APPLICANT:  Aftiran, Douglas
9      APPLICANT:  Rawlings, David
10     TITLE OF INVENTION:  HEMATOPOIETIC RESTRICTED TYROSINE KINASE
11     TITLE OF INVENTION:
12     NUMBER OF SEQUENCES:  7
13     CORRESPONDENCE ADDRESS:
14     ADDRESSEE:  FLEHR, HOBACH, TEST, ALBERTTTON & HERBERT
15     STREET:     4 Embarcadero Center, Suite 3400
16     CITY:       San Francisco
17     STATE:      California
18     COUNTRY:    USA
19     ZIP:        94111-4187
20     COMPUTER READABLE FORM:
21     MEDIUM TYPE: Floppy disk
22     COMPUTER:   IBM PC compatible
23     OPERATING SYSTEM:  PC-DOS/MS-DOS
24     SOFTWARE:   PatentIn Release #1.0, Version #1.25
25     CURRENT APPLICATION DATA:
26     APPLICATION NUMBER:  US/08/391,615
27     FILING DATE:
28     CLASSIFICATION:  435
29     PRIOR APPLICATION DATA:
30     APPLICATION NUMBER:  US 08/006,449
31     FILING DATE:       21-JAN-1993
32     ATTORNEY/AGENT INFORMATION:
33     NAME:              Rowland, Berttram I
34     REGISTRATION NUMBER:  20,015
35     REFERENCE/DOCKET NUMBER:  A-57583-1/BIR  UCLA 232-1
36     TELECOMMUNICATION INFORMATION:
37     TELEPHONE:         (415) 781-1989
38     TELEFAX:           (415) 398-3249
39     TELEX:             910 277299 FHT UR
40     INFORMATION FOR SEQ ID NO:  2:
41     SEQUENCE CHARACTERISTICS:
42     LENGTH:            659 amino acids
43     TYPE:              amino acid
44     TOPOLOGY:          linear
45     MOLECULE TYPE:     protein
46     US-08-391-615-2

```

Query Match	7.8%	Score 108.5;	DB 1;	Length 659;
Best Local Similarity	22.5%;	Pred. No. 0.16;		
Matches	52;	Conservative	41;	Mismatches 85;
				Indels 53;
				Gaps 12

[illegible]

QY 201 ALYTFEQNEGTSVVEEETLYVIEEDKDGWTRIRKNEDEEGYPTSYV 250
 ||| : | : : || : || | : ||| : || :
 Db 220 VALYDYMPMNANDLQLRGGEYFILEESNLP-WWRARDKNGEGYIPSNYI 269

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

Query Match	7.8%;	Score 108.5;	DB 3;	Length 1151;
Best Local Similarity	20.6%;	Pred. No. 0.33;		
Matches	58;	Conservative	57;	Mismatches 88;
			Indels	79;
			Gaps	14

QY	35	IKKKKGAAPPEFSNM--PPEORRRKLQO--KVDELNKEIKEMDOQRAITKRMVDYLKNP	90
Db	730	IKKKQEAIVAVYVYALKKEMWARKKDLASLRLOQLFADVE--DEETWIKEREIAASTN	786
QY	91	QMGPASIDHKL-----AEVSONIEKLRYETOKFPAIMAE-----VEGRIPARNEQ	136
Db	787	RGKDLIGVONLTKHQAOLQAEIAMEHPEPRICAATQKGAMWEEGHFAAEDVAKAKIHELNQ	846
QY	137	-----ARROSLGYVS-----QNPPTVNNC--ADQRESPD-----	163
Db	847	WEALKAKASQCRDLEBSLQAQYFADANAEASMMREKPIVGSYTDGKDESDIEALLK	906
QY	164	-----GSYTE--EQSQSEMKVLATPDDFDEDEEPLPAIGTKALYETEGON	209
Db	907	HEALMSDLAAGSSIQALREQAQSCQQAAPT--DDETKRELVL-----ALYDQKES	957
QY	210	EGTISVVEGETLYIEDKDGQWTRIRRNDEDECVYTSYVE	251
Db	958	PREVTMKRGDILLTINSTNKD-WKKVEVFN--DRQGFVAAYAK	997

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/051,982A
FILING DATE: 08-JUL-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: WILSON, MARY J.
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 1579-193
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4100
TELEFAX: (703) 816-4100
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 401 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: linear
US-09-051-982A-5

Query Match 7.6%; Score 106; DB 4; Length 401;
Best Local Similarity 28.6%; Pred. No. 0.14;
Matches 42; Conservative 19; Mismatches 60; Indels 26; Gaps 7;

QY 126 VEGRLPARNE-----QAROSGLVDSON-----PTVNNCAODRSPDGSTYE 168
DB 253 IGGSLPYRPPSTISQTSIQNONGPFYSQNPVSDTPPPVEEVPD-ESPPPPPP 311
QY 169 EQSOSEKMKVLTATDFDEDFDEPLA---IGTCKALYTFEGQNEGTISVEGETLYVI 224
DB 312 EDYEEBAAYV--EYSDPYAEDPPWAPRSYLEKVAIYIDYTKDEDELSPFGALITYI 369
QY 225 EEDKDGWTRIRRNDEEGYPTSYE 251
DB 370 KKN-DDGWYEGVM-GVYGLFPGNYVE 394

RESULT 11
PCT-US95-05008-8
Sequence 8, Application PC/TUS9505008
GENERAL INFORMATION:
APPLICANT: Sugen, Inc.
APPLICANT: 515 Galveston Drive
APPLICANT: Redwood City, California 94063-4720
APPLICANT: United States of America
APPLICANT: Wissenschaften E.V.
APPLICANT: Holgarten Str. 2
APPLICANT: Munchen 80539
APPLICANT: Germany
TITLE OF INVENTION: Novel Megakaryocytic Protein Tyrosine
TITLE OF INVENTION: Kinases
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05008
FILING DATE: 24-APR-1995

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/232,545
FILING DATE: 22-APR-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7683-074
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)790-9090
TELEFAX: (212)869-9741
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 659 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
PCT-US95-05008-8

Query Match 7.6%; Score 105.5; DB 5; Length 659;
Best Local Similarity 22.1%; Pred. No. 0.3;
Matches 51; Conservative 42; Mismatches 85; Indels 53; Gaps 12;

QY 43 PEDFSLPDEQRRKIQQAVDELKQEMQROAITR---MKDVIYLNKPMQMDPASL 98
DB 69 PE-KNPPE---KQIPRGE---SSEMQISLIEEPYPPFYVYDEGP----- 110
QY 99 DHKLAESONIEKLRVETQKFEAMLAEEVGRLLPARNEQARQ-----SGLV--DSQNP 150
DB 111 ---LVFSPTELR-----KRMHQKNVIRYNSDLVQKTHPCWIDQYLCSQYAK 160
QY 151 TVNNCAODRSPDGSTYEEQSOSEKMKVLTATDFDEDFDEPLP-----AIGTCK---- 200
DB 161 NAMGC-QILENNNGSLKPPSSHRKTKPLPPTPEEQILKPLPPEPAPAVSTSLKVV 219
QY 201 -ALYTFEGQNEGTISVEGETLYVEEDKGDGWTIRRNDEEGYPTSY 250
DB 220 VALYIMPMNANDLQIRKGEYFILESNLP-WMRARDXNGEGYIIPSNVY 269

RESULT 12
US-09-308-022-6
Sequence 6, Application US/09308022
Patent No. 6291654
GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA, ET AL.
TITLE OF INVENTION: C3 BINDING PROTEIN OF STREPTOCOCCUS
TITLE OF INVENTION: PNEUMONIAE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: METTING, RAASCH & GEBHARDT, P.A.
STREET: 119 No. 6291654th Fourth Street, Suite 203
CITY: Minneapolis
STATE: Minnesota
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/308,022
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/20586
FILING DATE: 12-NOV-1997
APPLICATION NUMBER: 60/029,444

```

?      ATTORNEY/AGENT INFORMATION:
?      NAME: Louis Myers
?      REGISTRATION NUMBER: 35,965
?      REFERENCE/DOCKET NUMBER: BGP-191
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: (617)227-7400
?      TELEFAX: (617)227-5941
?      INFORMATION FOR SEQ ID NO: 2:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 553 amino acids
?      TYPE: amino acid
?      TOPOLOGY: linear
?      MOLECULE TYPE: peptide
?      FRAGMENT TYPE: internal
?
US-08-475-894-2

Query Match          7.4%; Score 102.5; DB 1; Length 553;
Best Local Similarity 20.1%; Pred. No. 0.43;
Matches 66; Conservative 50; Mismatches 119; Indels 93; Gaps 14;

```

```

Query Match          7.4%: Score 102.5; DB 1; Length 553;
Best Local Similarity 20.1%; Pred. No. 0.43;
Matches    66; Conservative   50; Mismatches 119; Indels   93; Gaps   14.

QY      4 TVSDNSLSNRSNGEGCPDLPFGKSKSGKLMP-FIKKNKGATPEDFNSLNPPEQRKKLKQKV 62
       | : | : | I | I | : : | : | : | : | : | : | : | : | : | : | : |
Db      84 TRHDELDTISGEIITNR-----KEDGXMWEGGINRGRLGFDPNF-----VRIKEEMKK 134

QY      63 DEL-NKEIQK---EMDQDAITKMKDYLLKN-----POMDPAFLDLHK 101
       | | | | : : : : : : : : : : : : | | | | : | : |
Db     135 DPLINKAPEKRLHEVPSSNSLSLSEFTILRTNKGRERRRRRCVAARYSLPOND--ELEIK 192

QY     102 LAEVSQNIETKRIVETQKFEALAEVEGLPAR-NEQARROSGLYDSONPPVTNNCAODR 159
       : : : : : : : : : | : | : | : | : | : | : | : : : : : :
Db     193 VGDIIIEVGGE-VEEGWMEGVNLNCKTGMFPSNFIELSGSEDGLGISQDEQLSKSLRRT 250

QY     160 ESPDGSTIEESQSE-----MYLATDFEDDEPDDE-----190
       : : : : : : : : : | : | : | : | : | : | : | : : : : : :
Db     251 TGSSEDDGSSSTSEKANGTAVATAIOPKRVKGVGFIDIFDKPIKLRPRSIYEENDFL 310

QY     191 -----EPILAIGT-----CVALTFEQSNCGTSVYGEGLTYIE 225
       : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db     311 PVERTICKKLTATTATPATPDSSSTEMDSRTRKSVDYCVIPPEYQANDDELLIKEGDIVTLIN 370

QY     226 EDKGD-GMTRIIRNNEDEGGVPTSYEV 252
       : | | | | : | : | : : : :
Db     371 KDCIDVGMWBEELN-GRGQVFPDNFVKL 397

```

RESULT 14
 US-08-484-710-2
 ; Sequence 2, Application US/08484710
 ; Patent No. 5656438
 ; GENERAL INFORMATION:
 ; APPLICANT: Yen-Ming Hsu
 ; TITLE OF INVENTION: THE CAIP-LIKE GENE FAMILY
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESSEE:
 ; ADDRESSEE: LAHIVE & COCKFIELD
 ; STREET: 60 State Street, Suite 510
 ; City: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109-1875
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/484,710
 ; FILING DATE:
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Louis Myers
 ; REGISTRATION NUMBER: 35 965

C:Accession: J05261
 R:Tsujii, E.; Tsujii, Y.; Misumi, Y.; Fujita, A.; Sasaguri, M.; Ideishi, M.; Arakawa, K.
 Biochem. Biophys. Res. Commun. 229, 134-138, 1996
 A:Title: Molecular cloning of a novel rat salt-tolerant protein by functional complement
 A:Reference number: J05261; MUID:97112415
 A:Accession: J05261
 A:Molecule type: DNA
 A:Residues: 1-496 <TSU>
 A:Cross-references: DDBJ:D50557
 A:Note: The authors translated the codon GCA for residue 155 as Arg
 C:Comment: This protein is involved in a hormone-dependent transcriptional pathway.
 A:Gene: STP

Query Match 29.1%; Score 404.5; DB 2; Length 496;
 Best Local Similarity 47.3%; Pred. No. 5.5e-19;
 Matches 89; Conservative 29; Mismatches 53; Indels 17; Gaps 5;

QY 1 MKRTVSDNSLSNRGEGKPDILKFGGSKGKLMPTIKKNGATPEDFSNLPPEORRKLQ 59
 Db 291 MNRVPSDSSLGTP--DGRPELRAASSRSRAKRWPGKKNKTYVTEDFSLPEEQRRKLQ 348
 QY 60 QKVDPLNEIQEMDQRAITKMDVYLNKPQMGDPASLDHKLAEVSQNIETKRVETQKF 119
 Db 349 QOLEERNRELQKEEDQRAIKMKDYETKTPQMGDPASLEPRIAETLGINRIETNVEQKY 408
 QY 120 EAMLAIEVGR-LPARNEQARQSGLYDSQNPPTV-----NNCAQDRSPSGSTE 168
 Db 409 EAMLAIEATRLVLSNRGDSLRSRHTPPD--PPTTAPPDSSSSSSNSGSDNNESSEPPS 465
 QY 169 EOSOSEEM 176
 Db 466 EEGQDTPM 473

RESULT 3
 T15992
 hypothetical protein F09E10.8 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
 C:Accession: T15992
 R:Geisler, C.; Gattung, S.
 submitted to the EMBL Data Library, November 1995
 A:Description: The sequence of C. elegans cosmid F09E10.
 A:Reference number: Z18443
 A:Accession: T15992
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-554 <GEI>
 A:Cross-references: EMBL:U01749; NID:G1118144; PID:G1118152; PIDN:AAB52489.1; GSPDB:GN00
 A:Experimental source: strain Bristol N2; clone F09E10
 C:Genetics:
 A:Gene: CESP:F09E10.8
 A:Map position: X
 A:Introns: 107/2; 211/1; 274/1; 479/3; 532/2

Query Match 23.7%; Score 330; DB 2; Length 554;
 Best Local Similarity 29.6%; Pred. No. 4e-14;
 Matches 81; Conservative 51; Mismatches 104; Indels 38; Gaps 7;

QY 5 VSDNSLSNRGEGKPDILKFGGSKGKLMPTIKKNGATPEDFSNLPPEORRKLQKVD 64
 Db 287 MSKNGKGVARKQSMHQFFGGGTADK-----KTDSG----DYGLTPPQARRAKTIAGKISD 337
 QY 65 LNKTIQEMDQRAITKMDVYLNKPQMGDPASLDHKLAEVSQNIETKRVETQKFAMLA 124
 Db 338 LEKKDRATQSRGVSKMQAAYRENPKLGNSDCAQALQAGHEDALSNQIQKTKILD 397
 QY 125 EVEGRLLPARNEQARQSGLYDSQNPPTV-----NCAQDRSPSGSTEEOS 171
 Db 398 DVNAQLAGAGLSAIVSG---SDTPPSIRSVSSASGVTSKRYNTINDAKRTNGVGGGR 454

QY 172 QSEMKVLATDFD-----DEFDDE--EPILAIGTKALYTEGONEGISTIVEGE 219
 Db 455 EEPSSGSGSDSDPTPTINGHGHRDELYECSNPNPVLAGAIAOFADGADGTIRMEANE 514
 QY 220 TLVIEEDKDGQWTRIR-NEDEEGVPTSYVEV 252
 Db 515 KLMLEKDEGDQWTRIRKKNNSADGVPSSYLKV 548

RESULT 4
 T23452
 hypothetical protein K08E3.3a - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T23452
 R:McMurray, A.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19743
 A:Accession: T23452
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-783 <WIL>
 A:Cross-references: EMBL:Z81568; PIDN:CAB04591.1; GSPDB:GN00021; CESP:K08E3.3a
 A:Experimental source: clone K08E3
 C:Genetics:
 A:Gene: CESP:K08E3.3a
 A:Map position: 3
 A:Introns: 43/3; 99/3; 133/3; 186/3; 225/2; 266/3; 320/2; 355/3; 424/1; 451/2; 477/1;

Query Match 17.4%; Score 241.5; DB 2; Length 783;
 Best Local Similarity 23.9%; Pred. No. 3.1e-08;
 Matches 75; Conservative 56; Mismatches 108; Indels 75; Gaps 9;

QY 8 NSLSNRGEGKPDILKFGGK-----SKGRL--WPIKKNKA--TPEDFSN----- 48
 Db 472 DSRTNDSADGS-----GKRLKSSPSKNNIRNFILGILKEKADKPEASNNQDLMTYDK 526
 QY 49 -LPPQORRKLQKVDLKEIQEMDQRAITKMDVYLNKPQMGDPASLDHKLAEVSQ 107
 Db 527 SKPAHVRLSCLRSKIRDEKQLEQAIOGREGITRLOQAYTTPDHQGNPSACREPLISYAK 586
 QY 108 NIEKLVETQKFEAMLAIEVGRLLPARNEQA-----RRQGLYDSQNPPTVNNCAQ 157
 Db 587 KIEKLMQIDHNKKEFYVAMLEMSVEEGGRSPFGGRPTPDTTRMSGSSSTNQSSKTIEDVL 646
 QY 158 DRESPDGYTEQSQE-----SEMKVLATDFD 184
 Db 647 SCEAGNSSADSSKNILROLFTTPKRLISSPKTSKSTPTPLRRRAEISSPKILRSSFS 706
 QY 185 DEF-----DDEPPLPAIGTCALYTFEGQNGCTISVGEGLYVIEEDKDGQWTRIRN 238
 Db 707 GAIRKSLSTPDSVKETAVYATLAFEFAKSSAETWISTDGETLLVLEHDHGDGWTTRKK 766
 QY 239 EDEE-GYVPTSYVE 251
 Db 767 HNEESGFVPTSYLQ 780

RESULT 5
 T23456
 hypothetical protein K08E3.3b - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T23456
 R:McMurray, A.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19743
 A:Accession: T23456
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-785 <WIL>

F:500-550/Domain: SH3 homology <SH3A>
F:584-632/Domain: SH3 homology <SH3B>

Query Match 9.7%; Score 135; DB 2; Length 633;
Best Local Similarity 19.6%; Pred. No. 0.18; Mismatches 109; Indels 112; Gaps 14;
Matches 65; Conservative 45;

QY 2 KRTVSDNLSNSRGSGKPDLPKFGKSKGLMPFIKK--KGATPEPF----- 46
DB 254 KRLQANSYVK---QNKPSLNTA-----IFIKHNKLNKKEPQDFYVKKPSVMMHDDK 302
QY 47 ----SNLPEPQRRKKLQKVDLKN---ETQKEMDQDAITMKDYLKPNQMDPASLD 99
DB 303 FAVPSLVEEDIRIKLAENYNSLDQTKQNLKSLTLNKKIKHEMKTNEDINATKPY 362
QY 100 HKLAE-----VSQNIETKLVETQKFEAMLAEGVGRLPANR-----QARQSG 142
DB 363 DTLKRYLVNVSPTSHETLKQAEVQ-----IESIQNNVPEEYDUSTDNIDLSKTKKSG 417
QY 143 LYDS-----QNPPT-----VNCAQDRESPDG 164
DB 418 IFSKFNHNLVNDKSPSSGGSTGNGGPHLITSLFNTSRTRRLGSAPRNAGD---SDN 474
QY 165 SYTEQSQSESEKKVLTATPDEDFDEEPLPAIGTKALYTFEGQNGTISVEGETLYVI 224
DB 475 NSIRTTSTNNMKTKTQNSSD-----GKNKVLAYAVQKDDDETTTPGDKISLV 523
QY 225 EEDKGGDGTIRIRNED---EEGYPTSYVEV 252
DB 524 ARDQSGMTKI--NNDTGTGVLVPTTYIRI 552

RESULT 9
nephrocystin - human (fragment)
T08855
C:Species: Homo sapiens (man)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T08855
R:Hidestrand: F.; Otto, E.; Rensling, C.; Nothwang, H.G.; Vollmer, M.; Adolphs, J.; Hanu
Nature Genet. 17, 149-153, 1997
A:Title: A novel gene encoding an SH3 domain protein is mutated in nephronophthisis type
A:Reference number: 216499; MUID:97467724
A:Accession: T08855
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-731 <HIL>
A:Cross-references: EMBL:AF023674; NID:g2460116; PIDN:AAC51771.1; PID:g2460117
C:Genetics:
A:Gene: NPHP1
A:Map position: 2q13
A:Note: mutated in juvenile nephronophthisis 1
C:Keywords: kidney

Query Match 8.9%; Score 123.5; DB 2; Length 731;
Best Local Similarity 23.0%; Pred. No. 1.2;
Matches 51; Conservative 37; Mismatches 83; Indels 51; Gaps 8;
QY 53 ORRKLOQKVDLKNKEIOKEMDQDAI--TKMKDYLKPNQMDPASLDHKLAEVSONIE 110
DB 13 RRNELKQOYDLSLSEQ---LKEALEPKRKQHLY-----QKCIQKQKQID 55
QY 111 KLVETQKFEAMLAEGVGRLPANRQARQ-----SGL-----YDSQNP 149
DB 56 ENKVALQK---LSKADSAFVANYNQRKEEHTLLDKLQQLGLAVTISRENTTEGA 111
QY 150 PTYVNCADNDESPGTYTEBQSESEKKVLTATPDEDFDEEPLPAIGTKALYTFEGQNG 209
DB 112 PTEEBESESDESDSGEEDAEKEEENSHKMSGEEYIAYG-----DFTAQ 165
QY 210 EGTISVEGETLYVIEDKDGWTRIRNDEDEGYPTSYVE 251

DB 166 VGDLPFKGEILLVIEK-KPDGWWIAKDAKNGELVPRITYLE 206

RESULT 10

adaptor protein Cms - human
T13151
C:Species: Homo sapiens (man)
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
C:Accession: T13151; T08754
R:Kirsch, K.H.; Georgescu, M.M.; Ishimaru, S.; Hanafusa, H.
Proc. Natl. Acad. Sci. U.S.A. 96, 6211-6216, 1999
A:Title: Cms: An adaptor molecule involved in cytoskeletal rearrangements.
A:Reference number: 217608; MUID:99272673
A:Accession: T13151
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-639 <KIR>
A:Cross-references: EMBL:AF146277; NID:g4960046; PID:g4960047; PIDN:AAD34595.1
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, March 1999
A:Reference number: 216471
A:Accession: T08754
A:Molecule type: mRNA
A:Residues: 548-639 <MAM>
A:Cross-references: EMBL:AL050105
A:Experimental source: adult uterus; clone DKFP586H0519
C:Genetics:
A:Note: DKFP586H0519.1
C:Complex: homodimer
C:Function:
A:Description: probably functions as a scaffolding molecule with a specialized role 1
C:Keywords: coiled coil; homodimer

Query Match 8.5%; Score 118.5; DB 2; Length 639;
Best Local Similarity 21.0%; Pred. No. 2.1;
Matches 66; Conservative 49; Mismatches 77; Indels 123; Gaps 15;

QY 56 KKLQCK---VDELN-----KEIOKEMDQDAITKMKDYLKPNQMDPASLDH 101
DB 30 KKLQEGMIEGLNRRGMPFNPFKETIKETIEFD-----DSLTKRRHGNVASLVOR 84
QY 102 LAEV-----SONIEKLVETQK-----FEAMLAEGVGRLPAR-----NEQ 136
DB 85 ISTYGLPAGCIQPHQFTGNIKK---KTKRKQCKVLFE-YIPQNEDELKQVDIIDINE 140
QY 137 ARR--QSGIYSONPPTYVNCADNDESPDGYTEBQSESE----- 175
DB 141 VEEGWSGTLNKKLGLPFSNFEVKELEVTDDGETHAODDSETVLAGPTSPISLGNVSE 200
QY 176 -----MKVLATDFDEE----- 187
DB 201 ASGSVTPPKRTIGCGIFLFGKSVKLTTRTSSSTEERKEPKPLDLSLQPKTQSVET 260
QY 188 --DDEEPLPAIGTKALYTFEGQNGTISVVEGETLYVIEDKGD-GWTRIRNDEEGY 244
DB 261 KTDTEGKIKAKYECRTLAEGTNEDELTFKEGTIHLISKETGAGWRCGLN-GKGGV 319
QY 245 VPTSY-VEVC-LDKN 257
DB 320 FPDNFAVDINELDKD 334

RESULT 11

hypotheical protein Y116A8C.36 - Caenorhabditis elegans
T31504
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T31504
R:McMurray, A.
submitted to the EMBL Data Library, October 1999
A:Reference number: Z21041
A:Accession: T31504

C:Superfamily: protein-tyrosine kinase src protein kinase homology; SH2 homology; SH3
C:Keywords: ATP; autophosphorylation; blocked amino end; lipoprotein; myristylation; phos-
F:87-136/Domain: SH3 homology <SH3>
F:147-244/Domain: SH2 homology <SH2>
F:264-523/Domain: protein kinase homology <KIN>
F:272-280/Region: protein kinase ATP-binding motif
F:2/modified site: myristylated amino end (Gly) (in mature form) #status predicted
F:294/Active site: Lys #status predicted
F:415,526/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predicted

Query Match	8.38;	Score 115;	DB 1;	Length 532;
Best Local Similarity	35.14;	Prod No 3 0;		

QY 80 TKMKDYLYLKNQMGPASLDHKLAEVSONIEKLRYETQRFEAWLLEVEGRILPARNEDQARR 139
||| | : ||| | : : : :
Db 4 TKS-----PREGPRSRSLDIAEGSH-----QPFTSLA---SQTPSKSLDSHR 45

QY 140 QSGLYDSQNPPTVNNCAQDRESDDGSYTEEGSQSEMKVLTATDFDEFDDEE---PLP-A 195
 46 PSG-----QPEFGNCDF--TPFG-----GVNFSPTITSPDRTGPLAGG 81

QY 196 IGTCALYTFEGQNEGTSIVVEGETLYVIEDKGDGWTIRIRNDEECYPTSYV 250
: | ||| : : : | || : : : || :
Db 82 VITFVALDYESRTEFTDLSFRKGRLQIVANNTEGDMWLARSLSSGGQGYIPSNYV 136

RESULT 15
750995

150995

N;Alternate names: protein B7F18.140

C; Species: *Neurospora crassa*

C:\Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:\Accession: M50005

C;Accession: T50995

R. Schulte, U.; Alagn, V.; Honeisel, J.; Brandt, P.; Farthmann, B.; Holland, R.; Nyakatura submitted to the Protein Sequence Database, July 2000

A;Reference number: Z25286

A;Accession: T50995

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1119 <SCH>

A;CROSS-references: EMBL:AL389891; GSPDB:GN00116; NCSP:B7F18.140
A:Experimental source: PAC clone F7E18; strain O57A

A; Experimental source: BAC clone B/F18; strain OR/4A

C; Genetics:

A:Gene: NCSP:B/E18.140
A:Man position: 6

```
A;map position: 0
A:Introns: 66/3:
```

INSTRUCTIONS: 00/3; 123/2; 473/1

Query Match	Score	DB 2	Length
8.38%	115		1119

QY	Matches	Conservative	Mismatches	Indels	Gaps
196	20	19	17	6	2
IGNCKALYTFEGNEGTSVVEGELTYIEEDKGDGWRIRRN----	EDDEE	-GVP	TSY	249	

00 4 LGVKNALIDITPQEGEBLISEGDLIVLEKSDQEDDMWKAKKANAAADDEPVGILPNNX 63
 QY 250 VE 251

Search completed: April 7, 2002, 16:09:14
Job time: 192 sec

QY 213 ISVEGETLVYIEEDKGDGWTIRIRNEDEGYPTSYVEVCID 255
 DB 73 ISMAEGEDLSLMEEDKGDGWTIRVRKKEGEGGYPTSYLRTLTN 115

RESULT 2
 ID YB65_SCHPO STANDARD; PRT; 642 AA.

AC 009746;

DT 01-NOV-1995 (Rel. 32, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DE HYPOTHETICAL 72.2 KDA PROTEIN C12C2.05C IN CHROMOSOME II.

GN SPEC12C2.05C.

OS Schizosaccharomyces pombe (Fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;

OC NCB1_TaxID=4896;

OX NCB1_TaxID=4896;

RP SEQUENCE FROM N.A.

RC STRAIN=972;

RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;

RL Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.

CC -1- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG

CC BINDING DOMAIN.

CC -1- SIMILARITY: TO YEAST YH114W.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; Z54140; CA90818.1; -

DR HSSP; P29354; 1GFD.

DR InterPro: IPR002219; DAG_PE-bind.

DR InterPro: IPR001060; FCH.

DR InterPro: IPR001452; SH3.

DR Pfam; PF001130; DAG_PE-bind. 1.

DR Pfam; PF000611; FCH; 1.

DR PRINTS; PRO0008; DAGPEDOMAIN.

DR SMART; SM00109; C1; 1.

DR SMART; SM00055; FCH; 1.

DR SMART; SM00326; SH3; 2.

DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.

DR PROSITE; PS50081; DAG_PE_BIND_DOM_2; 1.

DR PROSITE; PS50002; SH3; 2.

KW Hypothetical protein; SH3 domain; Zinc; Phorbol-ester binding.

FT DOMAIN 397 447 PHORBOL-ESTER AND DAG BINDING.

FT DOMAIN 521 575 SH3 1.

FT DOMAIN 584 642 SH3 2.

SEQUENCE 642 AA; 72216 MW; E6770ED15BD73D3C CRC64;

Query Match 10.9%; Score 152; DB 1; Length 642;

Best Local Similarity 22.2%; Pred. No. 0.01; Mismatches 115; Indels 76; Gaps 12;

Matches 66; Conservative 47; Mismatches 115; Indels 76; Gaps 12;

QY 8 NSLSNRGEGKPDLPFGKSGKGLMPFKKNGATPEDFVSNLPFQRRKKIQQKVDLNLK 67

DB 312 NLNLLNLVHKSNDL---GKQKGLVSLDSQLEGLRVD-----PNSAQSPESSKKASINL 363

QY 68 EIQKEMDQDAITMKDYVILKNPQMGDPASLDHKLAEVSONIEKLRYETQKFEANLAEE 127

DB 364 EGKELMKV---ARIEDLEVR-----INKITSVANNLE---EGGRFHDF-KHVS 404

QY 128 GRLP-----ANNE---QARRSGLYDSQNPPTVNNCA 156

DB 405 FKLPTSCSYCRELIWGLSKRGCVCKNGCFKHARELLVPAACKNGEPEVADDAVDISV 464

QY 157 QRESPEGSGYTEEQSOSKMYLATDFD-----DEFDDEEPLPAIGTK-----ALY 203

DB 465 TATDFDASASSSNAYESRYNTYTDMDSSLYQTSLSNWKTEETTPADPAKVGQVLY 524

QY 204 TPEGONEGTISVEGETLVYIEEDKGDGWTIRIRNEDEGYPTSYVEVCIDLNK 259

DB 525 DTGHEGVITASGEQETFLLEPPDGSQWRY-KIDGIDGLIPASTYVKLNDELNTSYTLD 583

QY 260 GAKTYI 265

DB 584 GDSYV 589

RESULT 3

YH4_YEAST

ID YH4_YEAST STANDARD; PRT; 633 AA.

AC P38822;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE HYPOTHETICAL 71.2 KDA PROTEIN IN ERP5-ORC6 INTERGENIC REGION.

GN YH114W.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OC NCB1_TaxID=4932;

OX NCB1_TaxID=4932;

RP SEQUENCE FROM N.A.

RC STRAIN=S288C / AB972;

RA MEDLINE=94378003; PubMed=8091229;

RA Johnson M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,

RA Kucaha T., Hillier L., Jier M., Johnston L., Langston Y.,

RA Lacroix P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,

RA Nham N., Rifkin L., Riles L., St Peter H., Trevisan E., Vaughan K.,

RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,

RA Vaudin M.;

RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome

RT VIII.", 265:2077-2082(1994).

RL Science 265:2077-2082(1994).

CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.

CC -1- SIMILARITY: TO S.POMBE SPEC12C2.05C.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U00059; AAB6850.1; -

DR PIR; S48956; S48956.

DR SGD; S0001156; BZ21.

DR InterPro: IPR001060; FCH.

DR InterPro: IPR001452; SH3.

DR Pfam; PF00611; FCH; 1.

DR Pfam; PF00018; SH3; 2.

DR SMART; SM00055; FCH; 1.

DR SMART; SM00326; SH3; 2.

DR PROSITE; PS50002; SH3; 2.

KW Hypothetical protein; SH3 domain.

FT DOMAIN 493 555 SH3 1.

FT DOMAIN 577 633 SH3 2.

SEQUENCE 633 AA; 71171 MW; 5C73DAC69611B41 CRC64;

Query Match 9.7%; Score 135; DB 1; Length 633;

Best Local Similarity 19.6%; Pred. No. 0.12; Mismatches 109; Indels 112; Gaps 14;

Matches 65; Conservative 45; Mismatches 109; Indels 112; Gaps 14;


```

Db 64 EBRKAVLKSIIEOGKLSDDLRAQIEADNKTALDELTLPPYKPKRTKQAQIAHEHGIQPL 123
QY 102 ---LAEVSQNIETKLEVEQKFEAMLAEEGRLPARNQOAROSGLYDSQNPPTVNNCAQ 157
    ||| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 124 ADVLAEEQPDVE-----AAQGYL---NENIPAKALDQARALIMQFAE 167
QY 158 DRESFDGSTEEOSESEKVLATD-----FDDEFDDEPLPAIGTCALYTEEGON 209
    | | :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 168 DAEL-IGTLRDLKLMNEAEIHTQVEGKETEGEKFSDFPDRFVRAMPSPHRAVLAVLRGN 226
QY 210 EGTISV 215
    | | :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 227 EGVINI 232

RESULT 6
YHGF_NEIMA
ID YHGF_NEIMA STANDARD: PRT: 757 AA.
AC P57072.2
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN NMA0194.
GN NMA0194.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
NC NCBL_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=22491 / SEROGROUP A / SEROTYPE 4A.
RC MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Stimpson M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RA "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
CC -1 SIMILARITY: CONTAINS 1 'SI MOTIF' DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC
CC
CC EMBL, AL162752; CAB83508.1; -.
DR InterPro: IPR003029; SI.
DR Pfam: PF00575; SI. 1.
DR SMART: SM00316; SI. 1.
DR Hypothetical protein: RNA-binding. Complete proteome.
KW DOMAIN 640 709 SI MOTIF.
FT CONFLICT 265 266 WL -> CY (IN REF. 1).
SQ SEQUENCE 757 AA; 83161 MW; C7E9731BD4BC03F3 CRC64;

Query Match 8.4%; Score 116.5; DB 1; Length 757;
Best Local Similarity 23.1%; Pred. No. 2.2;
Matches 43; Conservative 37; Mismatches 67; Indels 39; Gaps 7;

```

```

Db 168 DAEL-IGTLRDLKLMNEAEIHAQVVGKETEGEKFSDFPDRFVRAMPSPHRAVLAVLRGN 226
QY 210 EGTISV 215
    | | :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:| :|:|
Db 227 EGVINI 232

RESULT 7
SH3_MOUSE
ID SH3_MOUSE STANDARD: PRT: 347 AA.
AC Q62421;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SH3-CONTAINING GRB2-LIKE PROTEIN 3 (SH3 DOMAIN PROTEIN 2C) (SH3P13).
GN SH3GL3 OR SH3D2C OR SH3D2C.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NC NCBL_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=98294438; PubMed=9630982;
RA Sparks A.B., Hoffman N.G., McConnell S.J., Fowlkes D.M., Kay B.K.;
RT "Cloning of ligand targets: systematic isolation of SH3
RT domain-containing proteins."
RL Nat. Biotechnol. 14:741-744(1996).
CC -1 FUNCTION: MAY PLAY A REGULATOR ROLE IN SYNAPTIC VESICLE RECYCLING
CC (BY SIMILARITY).
CC -1 SUBUNIT: INTERACTS WITH SYNAPTOJANIN AND DYNAMIN I.
CC -1 SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1 SIMILARITY: BELONGS TO THE ENDOPHTILIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC
CC
CC EMBL, U58887; AAC72268.1; -.
DR HSSP: P29355; 3SEW.
DR MGD: MGI:700011; Sh3d2c2.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00018; SH3. 1.
DR PRINTS: PR00452; SH3DOMAIN.
DR SMART: SM00326; SH3. 1.
DR ProSite: PS00002; SH3. 1.
KW SH3 domain; Coiled coil; Multigene family.
FT DOMAIN 180 201 COILED COIL (POTENTIAL).
FT DOMAIN 285 344 SH3.
FT DOMAIN 273 276 POLY-SER.
SQ SEQUENCE 347 AA; 38934 MW; A2174642F853B5EB CRC64;

Query Match 8.3%; Score 115.5; DB 1; Length 347;
Best Local Similarity 23.1%; Pred. No. 1;
Matches 56; Conservative 40; Mismatches 97; Indels 59; Gaps 11;

```


DR PIR, B34104; B34104.
 DR HSSP; P00523; 2PTR.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR000980; SH2.
 DR InterPro: IPR001452; SH3.
 DR InterPro: IPR001245; Tyr_kin.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00017; SH2; 1.
 DR Pfam: PF00018; SH3; 1.
 DR PRINTS: PR00109; TYRKINASE.
 DR PRINTS: PR00401; SH3DOMAIN.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR SMART; SM00252; SH2; 1.
 DR SMART; SM00326; SH3; 1.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS50002; SH2; 1.
 DR PROSITE; PS50002; SH3; 1.
 DR Transferrase; ATP-binding; Tyrosine-protein kinase; Phosphorylation;
 KW Myristate; SH3 domain; SH2 domain.
 FT INIT MET 0
 FT LIPID 1
 FT DOMAIN 79 140
 FT DOMAIN 146 243
 FT DOMAIN 265 518
 FT NP_BIND 271 279
 FT BINDING 293 293
 FT ACT_SITE 384 384
 FT MOD_RES 414 414
 SQ SEQUENCE 531 AA; 59605 MW; ED04AC5BE09C51B CRC64;

Query Match 8.3%; Score 115; DB 1; Length 531;
 Best Local Similarity 25.1%; Pred 1.8; Mismatches 61; Indels 46; Gaps 8;
 Matches 44; Conservative 24; Mismatches 61; Indels 46; Gaps 8;

QY 80 TKMKDVKLNKPMQDPASLDHKLAEVSONIEKLRYETOKFEAMLAEEGRLPARNQARR 139
 Db 3 TTSK-----PREGGPRSRSLDIAEGSH-----QFTSLSA---SQTPSKSLDSHR 44
 QY 140 QSGLDSONPPVNNACADRESPDSYTEGQSEEMKVLATDFDEPDEE---PLP-A 195
 Db 45 PSG-----QPRGNCDL--TPFG-----GVNFSDTTSPQRTGPLAG 80
 QY 196 ICTKALTYFEQNGSTIVVGEFTLYVIEDKGDGWFIRRNDEEGYPTSYV 250
 Db 81 VTFPALVDYERETEDLSFRKGERLQIVNNTEGDMWLARSLSGQTGIIPSNIV 135

RESULT 10
 YHGF_NEIIMB STANDARD; PRT; 757 AA.
 AC 051152;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN NMB0075.
 GN NMB0075.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B1940 / SEROGROUP B;
 RX MEDLINE=96236055; PubMed=8655518;
 RA Petering H., Hammerschmidt S., Frosch M., van Putten J.P.M.,
 RA Ison C.A., Robertson B.D.;
 RT "Genes associated with meningococcal capsule complex are also found
 in Neisseria gonorrhoeae";
 RL J. Bacteriol. 178:3342-3345 (1996).
 [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / SEROGROUP B;
 RX MEDLINE=20117575; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 RA Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 RA Mason T., Clecko A., Parksey D.S., Blair E., Ciltone H., Clark E.B.,
 RA Cotton M.D., Utterback T.R., Khouli H., Qin H., Vannatter J.,
 RA Gill J., Scariato V., Masigiani V., Pizzo M., Grand G., Sun L.,
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
 MC58.";
 RL Science 287:1809-1815 (2000).
 CC -1- SIMILARITY: CONTAINS 1 'S1 MOTIF' DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L09189; AAC37046.1; -
 CC EMBL; AE002366; AAF40542.1; -
 CC HSSP; P05055; ISRO.
 CC TIGR; NMB0075; -
 DR InterPro: IPR000110; Ribosomal_S1.
 DR InterPro: IPR003029; S1.
 DR Pfam; PF00575; S1; 1.
 DR SMART; SM00316; S1; 1.
 KW Hypothetical protein; RNA-binding; Complete proteome.
 FT DOMAIN 640 709
 FT CONFLICT 265 266 WL->CV (IN REF. 1).
 FT SEQUENCE 757 AA; 83161 MW; 7B6851EEB5ED66AE CRC64;

Query Match 8.2%; Score 114.5; DB 1; Length 757;
 Best Local Similarity 22.6%; Pred No. 2.9; Mismatches 68; Indels 39; Gaps 7;
 Matches 42; Conservative 37; Mismatches 68; Indels 39; Gaps 7;

QY 52 EORRRKLQOKVDE--LNKEIQENDORDATKMKDYLK-NPQMDPASP----- 98
 Db 64 EBRKAVVLKLSIEQGLSDSLRAQIEADNKTALIEDLYLPKPRRTKQIAREHGLQPL 123
 QY 99 -DHKLAEVSONIEKLRYETOKFEAMLAEEGRLPARNQARRQSGLDSONPPVNNCAQ 157
 Db 124 ADVLLAEQSDVE-----AAAQYL--NENVPDPAKALDGAARALIMEQFAE 167
 QY 158 DRESPDGYTEGQSEEMKVLATD-----FDDEPDEEPLPAIGCKALTYFEQNG 209
 Db 168 DAEI-IGTLRDLNMEAEIHAQVVECKETEGKFSDFPHRPREVKRIMSHRALAVLRGN 226

QY 210 EGTISV 215
 Db 227 EGVLINI 232

RESULT 11
 SH32_RAT
 ID SH32_RAT STANDARD; PRT; 248 AA.
 AC 035179;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SH3-CONTAINING GNB2-LIKE PROTEIN 2 (SH3 DOMAIN PROTEIN 2A) (ENDOPHILIN
 DE 1) (SH3P4) (FRAGMENT).
 GN SH3GL2 OR SH3P4.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;


```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Brain;
RX MEDLINE=97385143; PubMed=9238017;
RA Ringstad N., Nemoto Y., De Camilli P.;
RT "The SH3p1/SH3p8/SH3p13 protein family: binding partners for
RT synaptotagmin and dynamin via a Grb2-like Src homology 3 domain.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8569-8574(1997).
CC -1- FUNCTION: PLAYS A ROLE IN SYNAPTIC VESICLE RECYCLING, IN
CC PARTICULAR IN CLATHRIN-MEDIATED VESICLE ENDOCYTOSIS. EXHIBITS
CC LYSOPHOSPHATIDIC ACID ACYL TRANSFERASE ACTIVITY (LPAAT) (BY
CC SIMILARITY).
CC -1- SUBUNIT: INTERACTS WITH SYNAPTOJANIN AND DYNAMIN I (BY
CC SIMILARITY).
CC -1- TISSUE SPECIFICITY: BRAIN. EXPRESSED AT LOW LEVEL IN THE KIDNEY.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN
CC -1- SIMILARITY: BELONGS TO THE ENDOPHILIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF009603; AAC14883.1; -
CC DR HSSP: P29355; 3SEM.
CC DR InterPro: IPR000108; Neu_cyt_fact_2.
CC DR InterPro: IPR001452; SH3.
CC DR Pfam: PF00018; SH3; 1.
CC DR SMART: SM00326; SH3; 1.
CC DR PROSITE: PS50002; SH3; 1.
CC KW SH3 domain; Coiled coil; Multigene family.
CC FT NON_TER 1 1
CC FT DOMAIN 77 144 COILED COIL (POTENTIAL).
CC FT DOMAIN 186 245 SH3.
CC FT SEQUENCE 248 AA; 28371 MW; 50EC1ADCAS3EF1E1 CRC64;
SQ

Query Match 8.1%; Score 113; DB 1; Length 248;
Best Local Similarity 21.9%; Pred. No. 0.97; Mismatches 89; Indels 28; Gaps 7;
Matches 44; Conservative 40;

OY 55 RKLLQKVELINKIEMDQDAITKMKDVYLNKPDQMPASIDHKLAEVSNIEKLRV 114
DB 67 KKRROGKIPD-BELKQALEKFEDESKIEASSMFLNLEMDIEOVLSALVQAOLEHKK 124
OY 115 ETQKFEAMLAIEVGRLLPARNEDQAROSGLYDSQNPPTVNNCAODRESPDGSTEEOSES 174
DB 125 AVQILQOVVRLERIRIROSQSPRRE---YQPKRMSLFEPATGQNPNGCL---SHTG 177
OY 175 EMVNLATDPDFDEDEPRLPAIGTCKALYTFEGONESTISVSGETLYV---TEEDKGDG 231
DB 178 TPRPAGVQMD-----QP-----CCRALYDFEPENEGELGFKGDIITLLNIDENWYEG 226
OY 232 WTRIRNDEDEGYVTSYVEV 252
DB 227 ML-----HGSGFFPIYVEI 242

RESULT 12
SRC_RAT STANDARD; PRT; 535 AA.
ID O9MD9;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE SRC (EC 2.7.1.112) (P60-SRC)
DE (C-SRC).
GN SRC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCB1_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Testis;
RA Stockand J.D., Al-Khalil O., Spier B.J., Eaton D.C.;
RT "Rattus norvegicus proto-oncogene encoding tyrosine-protein kinase
RT pp60-c-src.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF130457; AAD24180.1; -
CC DR InterPro: IPR000719; Euk_pkinase.
CC DR InterPro: IPR000980; SH2.
CC DR InterPro: IPR001452; SH3.
CC DR InterPro: IPR001245; Tyr_kin.
CC DR Pfam: PF00069; pkinase; 1.
CC DR Pfam: PF00017; SH2; 1.
CC DR Pfam: PF00018; SH3; 1.
CC DR PRINTS: PR00109; TYRKINASE.
CC DR PRINTS: PR00401; SH2DOMAIN.
CC DR PRINTS: PR00452; SH3DOMAIN.
CC DR SMART: SM00252; SH2; 1.
CC DR SMART: SM00326; SH3; 1.
CC DR SMART: SM00219; TYRK; 1.
CC DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
CC DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC DR PROSITE: PS50001; SH2; 1.
CC DR PROSITE: PS50002; SH3; 1.
CC DR TRANSFERASE: Tyrosine-protein kinase; Proto-oncogene; Phosphorylation;
CC KW ATP-binding; Myristate; SH3 domain; SH2 domain.
CC FT INIT_MET 0 0
CC FT LIPID 1 1 MYRISTATE (BY SIMILARITY).
CC FT DOMAIN 83 144 SH3.
CC FT DOMAIN 150 247 SH2.
CC FT DOMAIN 269 522 PROTEIN KINASE.
CC FT NP_BIND 275 283 ATP (BY SIMILARITY).
CC FT BINDING 297 297 ATP (BY SIMILARITY).
CC FT ACT_SITE 386 388 BY SIMILARITY.
CC FT MOD_RES 419 419 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC FT MOD_RES 529 529 PHOSPHORYLATION (BY SIMILARITY).
CC FT SEQUENCE 535 AA; 59946 MW; AD083DD2357890DC CRC64;
SQ

Query Match 7.9%; Score 110.5; DB 1; Length 535;
Best Local Similarity 23.6%; Pred. No. 3.4; Mismatches 67; Indels 5; Gaps 2;
Matches 29; Conservative 22;

OY 131 PARNEQARRQSGLYDSQNPPTVNNCAODRESPDGSTEEOSEEMKVLATDPDEFDE 190
DB 19 PAEN--VHGAGAPRPASTGPKSPASAOCHGRPNMAFVPPAAEPKLGFGNSDPTVSPQ 76
OY 191 EPLPAIG---TCKALYTFEGONESTISVSGETLYVIEEDKGDGWTIRRNDEDEGYPT 247
DB 77 RAGLGGVTVTFVALYDESRTETDLSEFKGERLQIYVNNPGLWLAHSLSTGQTYLPS 136
OY 248 SYV 250
DB 137 NYV 139

```

```

RESULT 13
MYSC_CAEEL STANDARD: PRT: 1947 AA.
AC P12845;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN HEAVY CHAIN C (MHC C).
GN MYO-2
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=89178677; PubMed=2926820;
RA Dibb N.J., Maruyama I.N., Krause M., Karn J.;
RT "Sequence analysis of the complete Caenorhabditis elegans myosin
RL heavy chain gene family."
RL J. Mol. Biol. 205:603-613(1989).
RN [2]
RP SEQUENCE OF 1-21; 112-371 AND 1501-1772 FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=85201409; PubMed=388374;
RA Karn J., Dibb N.J., Miller D.M.;
RT "Cloning nematode myosin genes."
RL Cell Muscle Motil. 6:185-237(1985).
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATOR LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.
CC -1- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE PHARYNGEAL MUSCLE.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEMOROSIN (LMM) AND 1 HEAVY MEMOROSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN C.
CC ELEGANS.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X08066; CAA30855.1; -
DR EMBL: M37233; AAA28121.1; -
DR EMBL: M37235; AAA28122.1; -
DR EMBL: M37236; AAA28123.1; -
DR PIR: S05697; S05697.
DR HSSP: P08799; IMND.
DR InterPro: IPR002928; Myosin_tail.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head.1.
DR Pfam: PF01576; myosin_tail.1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODOM: PD000355; myosin_head.1.
DR SMART: SM00242; MYSC.1.
DR SMART: SMO0242; MYSC.1.
KW ATP-binding; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Multigene family.
FT DOMAIN 1 855 MYOSIN HEAD-LIKE.

```

```

FT DOMAIN 856 1947 RODLIKE TAIL (S2 AND LMM DOMAINS).
FT DOMAIN 856 1947 COILED COIL (POTENTIAL).
FT NP_BIND 174 181 ATP (BY SIMILARITY).
FT DOMAIN 668 690 ACTIN-BINDING.
FT DOMAIN 773 787 ACTIN-BINDING.
FT MOD_RES 125 125 METHYLATION (TR1-) (POTENTIAL).
FT MOD_RES 708 708 ALKYLATION (SH-1).
FT MOD_RES 718 718 ALKYLATION (SH-2).
FT CONFLICT 132 132 E -> D (TN REF. 2).
FT CONFLICT 137 137 M -> I (TN REF. 2).
SQ SEQUENCE 1947 AA; 223009 MW; 6D65AE9BD013627 CRC64;

Query Match
Best local similarity 7.9%; Score 110; DB 1; Length 1947;
Matches 56; Conservative 48; Mismatches 106; Indels 52; Gaps 10;

QY 26 KSKGLMPFIKKNK-----GATPEDFSLLPPEQRKKLQOKVDELNKEIQEND 74
DB 847 KLYGKVPPLVNSGRTBAQYKELQETVALTKD-TVQGEKKRQLQGAERLNKETFADLLA 905
QY 75 QRDATL-KKKDYLLKNPQGD-PASLDHKLAEVSQIE-----K 111
DB 906 QLEASGSTRVEEEMRTANNEQKVALEGLADASKLEVEEARAVEINKOKLVEACAD 965
QY 112 LRVETQFPFAMLAIEVGRLLPARNEQAR-----ROSLGVDSQNPPTVNNCAODRESPDGS 165
DB 966 LKNCQDVDSLKKVAEKNNAKHQIRALODEKRO-----QDEIKSLNKRKNQESQNK 1021
QY 166 YTEESQSEEMKVLADPDEF-----DDEEPLPAITGCKA-LYPEGQNEGTSIVEGE 219
DB 1022 LVEDLQAEENQLAANKLAKLMQSLSEDEQWTERKRRNRADMDKKRAEGELKTAQ-E 1080
QY 220 TLVVEEDKGDGTRRRNDE 241
DB 1081 TLEELNKSXSDAENLARRETE 1102

RESULT 14
HSL_MOUSE STANDARD: PRT: 486 AA.
AC P49710;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HEMATOPOIETIC LINEAGE CELL SPECIFIC PROTEIN (HEMATOPOIETIC CELL-
DE SPECIFIC LYN SUBSTRATE 1) (LCKBPI).
GN HCS1 OR HSL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95217198; PubMed=7535527;
RA Kitemura D., Kaneko H., Tanluchi I., Yamamura K.I., Watanabe T.;
RT "Molecular cloning and characterization of mouse HSL."
RL Biochem. Biophys. Res. Commun. 208:1137-1146(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/C;
RX MEDLINE=97098949; PubMed=8943564;
RA Takemoto Y., Sato M., Furuta M., Hashimoto Y.;
RT "Distinct binding patterns of HSL to the Src SH2 and SH3 domains
RT reflect possible mechanisms of recruitment and activation of
RT downstream molecules."
RL Int. Immunol. 8:1699-1705(1996).
CC -1- FUNCTION: SUBSTRATE OF THE ANTIGEN RECEPTOR-COUPLED TYROSINE
CC KINASE. PLAYS A ROLE IN ANTIGEN RECEPTOR SIGNALING FOR BOTH CLONAL
CC EXPANSION AND DELETION IN LYMPHOID CELLS. DIRECTLY ASSOCIATES WITH
CC HAX-1, THROUGH BINDING TO ITS C-TERMINAL REGION. MAY ALSO BE
CC INVOLVED IN THE REGULATION OF GENE EXPRESSION (BY SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH THE SH2 AND SH3 DOMAINS OF LCK.

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2002, 16:08:16 ; Search time 40.44 Seconds
(without alignments)
320.545 Million cell updates/sec

Title: US-09-925-122A-3
Perfect score: 910
Sequence: 1 MKDVYKTPQMGDPASLEPQ.....RRKEGEGYPTSYLRVTLN 175

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_1101.*
1: /SIDSB/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSB/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSB/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSB/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSB/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSB/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSB/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSB/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSB/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSB/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSB/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSB/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSB/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSB/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSB/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSB/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSB/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSB/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSB/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSB/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSB/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSB/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	908	99.8	175	20	AAV24923
2	538	59.1	211	20	AAV40582
3	538	59.1	211	20	AAW85101
4	538	59.1	212	20	AAW92389
5	533	58.6	212	15	AAW53543
6	450	49.5	217	22	AAE04186
7	450	49.5	367	22	AAE04187
8	449	49.3	547	22	AAW61130
9	432.5	47.5	265	20	AAV24922
10	432.5	47.5	537	22	AAW92528
11	432.5	47.5	592	22	AAW94690

12	131.5	14.5	370	21	AAV94428	Human PRO1431 (UNQ)
13	131.5	14.5	370	22	AAW66177	Protein of the inv
14	123.5	13.6	377	17	AAW05411	Human H74 protein.
15	123.5	13.6	486	21	AAW43338	Human ORFX ORF3102
16	123.5	13.6	486	21	AAW54041	Protein encoded by
17	119	13.1	659	17	AAW06708	Mouse haematopoiet
18	118	13.0	180	17	AAW05400	Human clone 34 pro
19	118	13.0	733	22	AAW39073	Human polypeptide
20	118	13.0	733	22	AAW40859	Human polypeptide
21	117	12.9	659	17	AAW94534	BRK tyrosine kinase
22	116.5	12.8	486	22	AAW24503	Human PAC amino ac
23	114.5	12.6	739	19	AAW61532	Mouse Fas-binding
24	114.5	12.6	739	22	AAW69150	Mouse daxx protein
25	110	12.1	648	21	AAW95048	Candida albicans p
26	109	12.0	424	21	AAW41838	Human ORFX ORF1602
27	109	12.0	424	21	AAW27227	Human EXMAD-5 SEQ
28	108	11.9	251	21	AAW44450	Mutant chicken c-S
29	108	11.9	533	14	AAW39705	Chicken pp60 c-src
30	108	11.9	533	21	AAW44447	Wild-type chicken
31	108	11.9	533	21	AAW44449	Mutant chicken c-S
32	108	11.9	533	21	AAW44451	Mutant chicken c-S
33	108	11.9	533	22	AAW84661	Amino acid sequenc
34	107.5	11.8	211	22	AAW36685	Mammalian two-hydr
35	106	11.6	442	17	AAW06709	Human haematopoiet
36	105	11.5	620	17	AAW94535	ITK tyrosine kinase
37	104.5	11.5	496	20	AAW29668	Human src-family k
38	104	11.4	536	14	AAW39706	Human pp60 c-src p
39	102.5	11.3	146	12	AAW15156	Wild-type human c-
40	101.5	11.2	450	21	AAW44448	PKA substrate, Csk
41	101.5	11.2	450	21	AAW44448	Amino acid sequenc
42	101.5	11.2	450	22	AAW84662	Mammalian two-hydr
43	100.5	11.0	214	22	AAW36681	Amino acid sequenc
44	100	11.0	397	20	AAW28292	Amino acid sequenc
45	100	11.0	416	20	AAW28291	

ALIGNMENTS

RESULT	1
AAV24923	
ID	AAV24923 standard; Protein; 175 AA.
AC	
XX	AAV24923;
DT	27-AUG-1999 (first entry)
XX	
DE	Human SH3-containing protein 2.
XX	
KW	Human SH3 containing protein; HS3C-1; HS3C-2; Src homology 3 domain;
KW	diagnosis; cancer; immune disorder; development disorder; leukemia;
KW	immunoflammatory condition; rheumatoid arthritis; ulcerative colitis;
KW	osteoarthritis; Gaucher's disease; adenocarcinoma; Lymphoma; melanoma;
KW	sarcoma; AIDS; allergy; asthma; irritable bowel syndrome; pancreatitis;
KW	multiple sclerosis; osteoarthritis; haemodialysis; infection; trauma;
XX	anaemia; epilepsy; congenital glaucoma.
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
FT	Misc-difference 65
FT	/label= unknown
FT	/note= "encoded by GAN"
XX	
XX	US5916753-A.
XX	
XX	29-JUN-1999.
PD	
XX	
XX	13-NOV-1997; 97US-0970133.
PF	
XX	
XX	13-NOV-1997; 97US-0970133.
PR	
XX	
XX	(INCY-) INCYTE PHARM INC.

XX Bandman O, Guegler KJ, Lal P;
 XX WPI: 1999-394206/33.
 DR N-PSDB; AAX83629.
 XX
 PT New Src homology 3 domain containing proteins useful for the
 PT diagnosis, treatment or prevention of cancer and immune or
 PT development disorders
 XX
 PS Claim 8; Fig 2; 32pp; English.
 XX
 CC The present sequence represents human Src homology 3 domain (SH3)
 CC containing protein 2, designated HS3C-2. HS3C proteins can be used
 CC for the diagnosis, treatment or prevention of cancer and immune or
 CC development disorders. HS3C-1 is particularly expressed in prostate
 CC tissues associated with prostate tumours and HS3C-2 with inflammatory
 CC conditions such as rheumatoid arthritis, ulcerative colitis,
 CC osteoarthritis and Gaucher's disease. A vector expressing the complement
 CC of the polynucleotide encoding HS3C-1 can be administered to a subject
 CC to prevent or treat cancers including adenocarcinoma, leukemia, lymphoma,
 CC melanoma, sarcoma and especially cancers of the bladder, kidney, heart,
 CC lung, adrenal gland, skin, spleen, liver, ovary, pancreas, thyroid and
 CC uterus. An immune disorder such as AIDS, allergies, asthma, irritable
 CC bowel syndrome, multiple sclerosis, pancreatitis and osteoarthritis can
 CC also be treated along with complications of cancer, haemodialysis, viral,
 CC bacterial, fungal, and parasitic infections and trauma. A vector
 CC expressing the complement of the polynucleotide encoding HS3C-2 can also
 CC be administered to a subject to prevent or treat cancers and immune
 CC disorders as well as developmental disorders such as anaemia, epilepsy,
 CC and congenital glaucoma. The expression vectors which encode HS3C can be
 CC used to deliver nucleotide sequences to targeted organ, tissue or cell
 CC populations and antisense polynucleotides to treat conditions associated
 CC with overexpression of HS3C by blocking transcription of the mRNA,
 CC modulating HS3C activity or regulating the gene function.
 CC
 XX
 SQ Sequence 175 AA:
 QY
 Query Match 99.8%; Score 908; DB 20; Length 175;
 Best Local Similarity 100.0%; Pred. NO. 2e-75;
 Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKDYKTPQMGDPASLEPOIAETLSNTERLKLVEQKYEAWEAESEVLSNRGSLSRH 60
 Db 1 mkdyktpmgdpaslepoiaetlsnlerlklevkyeawaeesvlsnrgslsrh 60
 QY 61 ARPPPPASAPPPSSNSASQDTKESSEPPSESDPTPIYTFEDDFEERTSPIGHCV 120
 Db 61 arppppasapppssnsasqdtkesseppseesdptpiytfeddfеертспіghcv 120
 QY 121 AIYHFGESSEGTISMAEGEDLSLMEEDKGDGWTVRVRKREGGEGYVPTSYLRVTLN 175
 Db 121 aiyhfgessegtsmaegedlsimeedkgdgwtrvrkrreggegyvptsylyrvtn 175
 RESULT 2
 AAY40582
 ID AAY40582 standard; protein: 211 AA.
 XX
 AC AAY40582;
 XX
 DT 03-DEC-1999 (first entry)
 XX
 DE Partial amino acid sequence of TR-interacting protein S410a.
 XX
 KW J11 protein; thyroid hormone receptor; trap assay; therapeutic;
 KW nuclear hormone receptor protein; TR-interacting protein; human;
 KW thyroid disorder.
 XX
 OS Homo sapiens.
 XX
 PN US5962256-A.

XX
 PD 05-OCT-1999.
 XX
 PF 06-JUN-1995; 95US-0471613.
 XX
 PR 04-APR-1994; 94US-0222719.
 PR 30-OCT-1992; 92US-0969136.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 XX Lee JW, Moore DD;
 PI
 DR WPI: 1999-571268/48.
 DR N-PSDB; AA207568.
 XX
 PT Purified DNA comprising a sequence encoding a protein which
 PT specifically interacts with a thyroid hormone receptor -
 XX
 PS Disclosure; Fig 14; 68pp; English.
 CC
 CC The invention provides a J11 protein which specifically interacts with a
 CC thyroid hormone receptor in an in vivo trap assay. The J11 protein can
 CC be recombinantly produced by standard recombinant methodology. The J11
 CC protein is used in a trap assay for determining whether a test protein
 CC is capable of interacting with a nuclear hormone receptor protein, and
 CC may also be used as a therapeutic peptide for treating thyroid
 CC disorders. Sequences AA140572-596 represent partial amino acid sequences
 CC of thyroid hormone receptor (TR)-interacting proteins.
 CC
 XX
 SQ Sequence 211 AA:
 QY
 Query Match 59.1%; Score 538; DB 20; Length 211;
 Best Local Similarity 99.0%; Pred. NO. 1.7e-41;
 Matches 102; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 73 DSSNSASQDTKESSEPPSESDPTPIYTFEDDFEERTSPIGHCVAIYHFGESSEGT 132
 Db 13 nssnsasqdtkesseppseesdptpiytfeddfеертспіghcvaiyhfgesseg 72
 QY 133 ISMAEGEDLSLMEEDKGDGWTVRVRKREGGEGYVPTSYLRVTLN 175
 Db 73 ismaegedlsimeedkgdgwtrvrkrreggegyvptsylyrvtn 115
 RESULT 3
 AAW85101
 ID AAW85101 standard; Protein: 211 AA.
 XX
 AC AAW85101;
 XX
 DT 09-FEB-1999 (first entry)
 XX
 DE Thyroid hormone receptor-Interacting protein S410a partial sequence.
 XX
 KW Thyroid hormone receptor-Interacting protein; S410a;
 KW nuclear hormone receptor protein; screen.
 XX
 OS Homo sapiens.
 XX
 PN US5846711-A.
 XX
 PD 08-DEC-1998.
 XX
 PF 04-APR-1994; 94US-0222719.
 XX
 PR 04-APR-1994; 94US-0222719.
 PR 30-OCT-1992; 92US-0969136.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 XX Lee JW, Moore DD;
 PI

DR	WPI: 1999-059040/005.
DR	N-PSDB; AAV82591L.
XX	
FT	Screening assay for nuclear hormone receptor modulators - using
PT	cells containing reporter gene construct
XX	
PS	Disclosure; Fig 14; 69pp; English.
XX	
CC	The present sequence represents thyroid hormone receptor-interacting
CC	protein S410a. The protein was identified using the method of the
CC	Invention. The method is used to determine if a test protein is capable
CC	of interacting with a nuclear hormone receptor protein in a
CC	ligand-dependent manner. The method is used especially to screen for
CC	proteins that interact with thyroid hormone receptors in a
CC	ligand-dependent or ligand-sensitive manner.
XX	
SO	Sequence 211 AA;
OY	Query Match 59.1%; Score 538; DB 20; Length 211; Best Local Similarity 99.0%; Pred. No. 1.7e-41;
Db	Matches 102; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY	73 DSSNSASQDKRESSEPPSESQDTPIYTEDEDFEEBPSPICHCVAIHFESSSECT 132 : : : : : : : : : 13 nssnsaaqdktesseepseeqqdtplytlededfeepsfshcvaiahfgssegt 72
OY	133 ISMAEGEDLSLMEEKDGDGWTFRRKEGEGVPTSYLRVTLN 175 : : : : : : : : : 73 lsmaegedlsimeedkgdgwtfrrrkgeggvptsyrlrvtn 115
Db	
RESULT	4
ID	AAW923389
XX	AAW923389 standard; Protein; 212 AA.
AC	AAW923389;
XX	
DT	21-APR-1999 (first entry)
XX	
DE	Human TR-interacting protein S410a.
XX	
KM	Thyroid hormone receptor-interacting protein; TR-interacting protein;
KW	JLI; human; interaction trap assay; treatment; thyroid disorder; S410a;
KM	hyperthyroidism.
XX	
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	Protein 1..212
FT	/note= "Partial coding sequence, no start or stop
FT	codons given"
FT	Misc-difference 116
FT	/note= "in frame stop codon encoded by TGA"
XX	
PN	US5866686-A.
XX	
PD	
XX	02-FEB-1999.
XX	
PF	06-JUN-1995; 95US-0470925.
XX	
PR	04-APR-1994; 94US-0222719.
PR	30-OCT-1992; 92US-0969136.
PR	06-JUN-1995; 95US-0470925.
XX	
PA	(GEHO) GEN HOSPITAL CORP.
XX	
PI	Lee JW, Moore DD;
XX	
DR	WPI: 1999-142013/12.
DR	N-PSDB; AAX01880.
XX	
PT	JLI protein - that interacts with thyroid hormone receptor

XX		Disclosure; Fig 14; 67pp; English.
PS		
XX		This sequence represents a human thyroid hormone receptor (TR-interacting
CC		protein) which is found to interact with a novel human JLI protein in an
CC		in vivo interaction trap assay. The JLI protein is potentially useful for
CC		the treatment of hyperthyroidism or thyroid disorders.
CC		
XX		
SO	Sequence	212 AA;
OY	Query Match	59.1%; Score 538; DB 20; Length 212;
Dd	Best Local Similarity	99.0%; Pred. No. 1.7e-41;
	Matches 102; Conservative	1; Mismatches 0; Indels 0; Gaps 0
OY	73 DSSNSASODRKESSEPPSESDPPIRYEPEDDEEPEPTSGHCVAIHPESSECT	132
	:	
Dd	13 nssnsaqdkkesseepseeqdprlyetedeleeptspgncvalynfegssegt	72
OY	133 ISMAEGEDLSIMEBDKGDGWTFRARKKEGGGYVPTYSLRVTLN	175
Dd	73 ismaeedlsimeedkgdgwrtvrtrkgegyvptyslrvtlm	115
RESULT	5	
AAR53543		
ID	AAR53543 standard; protein; 212 AA.	
XX	AC	AAR53543;
XX		
DI	21-FEB-1995	(first entry)
XX		
DE	Thyroid hormone receptor-interacting protein - S410a.	
XX		
KW	nuclear thyroid hormone interacting proteins; TR; JLI; JI2; S410a;	
KM	transcriptional coactivator; treatment; diagnosis; SH3 domains;	
KW	thyroid related disorders; modulation; thyroid hormone receptor;	
KM	nuclear hormone receptor; isolation.	
XX		
OS	Chimeric Homo sapiens.	
OS	Chimeric Bacterial sp.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 116	/note= "stop codon encoded by TGA"
FT		
PN	WO9410338-A.	
XX		
PD	11-MAY-1994.	
XX		
PF	29-OCT-1993;	93MO-US10443.
XX		
PR	30-OCT-1992;	92US-0969136.
XX		
PA	(GEHO) GEN HOSPITAL CORP.	
XX		
PI	Lee JW, Moore DD;	
XX		
DR	WPI; 1994-199808/24.	
XX		
PT	Nuclear hormone receptor interacting polypeptides, esp. thyroid	
PT	hormone=interacting proteins (TRs) - for identifying proteins	
PT	useful in treatment and diagnosis of thyroid related disorders by	
PT	inoculating thyroid hormone receptor activity	
XX		
PS	Claim 21; Page 50-51; 105pp; English.	
XX		
CC	This sequence shows the partial amino acid sequence of S410a	
CC	(containing a SH3 domain), a thyroid hormone (TR) interacting protein.	
CC	TR-interacting proteins physically associate with thyroid hormone	
CC	receptor. Nearly all the fusion cDNAs showed very strong dependence	
CC	on hormone activation. The proteins can be used in an in vivo trap	
CC	system for the isolation of proteins which associate with any nuclear	

CC hormone receptor. The proteins and Abs may be used to treat or diagnose
CC thyroid disorders, and to modulate thyroid hormone receptor activity.
XX
SQ Sequence 212 AA;

Query Match 58.6%; Score 533; DB 15; Length 212;
Best Local Similarity 98.1%; Pred. No. 4.8e-41;
Matches 101; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY DSSNSASODRKESSEPPSESDPTIYTFEDDEFEPPSPICGCAVIAHFECSSEGT 132
Db :|||||
13 nssnsasqdkesceesepesqclpyletdeefeeptspighcvaahfegsseg 72

QY 133 ISMAEGEDLSLMEEDKGDGWTVRRRKEGEGYVPSYLRVTLN 175
Db :|||||
73 ismaegedlsimeedkgdgwtvrtrrrkeggyvpsylrvtl 115

RESULT 6
AAE04186
ID AAE04186 standard; Protein: 217 AA.
XX
AC AAE04186;
XX
DT 09-AUG-2001 (first entry)
XX
DE Human gene 10 encoded secreted protein fragment, SEQ ID NO:178.

Human: secreted protein: proliferative disorder; cancer; tumour; asthma;
foetal abnormality; developmental abnormality; haematopoietic disorder;
immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
inflammation; neurological disorder; Alzheimer's disease; food additive;
angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
pregnancy-related disorder; endocrine disorder; infection; wound healing;
cell culture; chemotaxis; vulnery; binding partner Identification;
gene therapy; chromosome 1.

XX Homo sapiens.
OS
XX
XX WO200134643-A1.
PN
XX
XX 17-MAY-2001.
PD
XX
XX 08-NOV-2000; 2000WO-US30629.
PF
XX
XX 12-NOV-1999; 99US-0164825.
PR
XX 03-AUG-2000; 2000US-0222904.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Ruben SM, Komatsoulis GA, Soppet DR, Shly;
PI
XX
XX WPI; 2001-374441/39.
DR

Nucleic acids encoding 24 human secreted polypeptides, useful for
preventing, diagnosing and/or treating e.g. Gaucher's disease,
Alzheimer's disease, scintar syndrome, Creutzfeldt-Jacob disease,
diabetes mellitus and multiple sclerosis -
PT
XX
XX
PS Disclosure; Page 32; 532pp; English.

CC AAD08404-AAD08478 represent cDNAs corresponding to 24 human secreted
CC protein genes, and AAE04100-AAE004170 represent the proteins they encode.
CC AAE04172-AAE04197 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing, treating,
CC or ameliorating medical conditions, e.g., by protein or gene therapy.
CC Pathological conditions can be diagnosed by determining the amount of the
CC new protein in a sample or by determining the presence of mutations in
CC the new genes. Specific uses are described for each of the 24 genes,
CC based on the tissues in which they are most highly expressed, and include

CC developing products for the diagnosis or treatment of proliferative
CC disorders, cancer, tumours, foetal and developmental abnormalities,
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
CC angiogenic disorders, kidney disorders, gastrointestinal disorders,
CC pregnancy-related disorders, endocrine disorders, and infections. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues, to identify their cognate ligands or binding
CC partners, and in chemotaxis, and can be used as a food additive or
CC preservative to modify storage properties. Antibodies specific for a
CC protein of the invention can be used in alleviating symptoms associated
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The
CC present sequence represents a human secreted protein fragment referred
CC to in the disclosure of the invention.

XX
SQ Sequence 217 AA;

Query Match 49.5%; Score 450; DB 22; Length 217;
Best Local Similarity 50.3%; Pred. No. 1.9e-33;
Matches 88; Conservative 30; Mismatches 49; Indels 8; Gaps 4;

QY 1 MKDYVEKTPQMDPASPSEPOIAETLSNTERKLEQVKEAMLAESRYLSNRGDSLSRH 60
Db :|||||
43 mkdyeknpqmgdpaspipklatennidrlrmeahkneawlsevgklt-99gtd--trh 99

QY 61 ARPPXPAPSDSSNSASODRKESSEPPSESDPTIYTFEDDE-DEPTSPICG 119
Db :|||||
100 ssdnhlvtgreespegyltdangevypqgnh----hnefdedfeddpipaignhc 155

QY 120 VAIYFECSSEGTSMAGEDLSLMEEDKGDGWTVRRRKEGEGYVPSYLRVTL 174
Db :|||||
156 kaiyfdghegtlamkegevllyleedkgdgwttrrrngeegypvpsyltdvlt 210

RESULT 7
AAE04187
ID AAE04187 standard; Protein: 367 AA.
XX
XX
XX AAE04187;
AC
XX
XX 09-AUG-2001 (first entry)
DT
XX
XX Human gene 10 encoded secreted protein fragment, SEQ ID NO:179.
DE
XX
XX Human: secreted protein: proliferative disorder; cancer; tumour; asthma;
foetal abnormality; developmental abnormality; haematopoietic disorder;
immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
inflammation; neurological disorder; Alzheimer's disease; food additive;
angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
pregnancy-related disorder; endocrine disorder; infection; wound healing;
cell culture; chemotaxis; vulnery; binding partner Identification;
gene therapy; chromosome 1.

XX Homo sapiens.
OS
XX
XX WO200134643-A1.
PN
XX
XX 17-MAY-2001.
PD
XX
XX 08-NOV-2000; 2000WO-US30629.
PF
XX
XX 12-NOV-1999; 99US-0164825.
PR
XX 03-AUG-2000; 2000US-0222904.
XX

PA (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Komatsoulis GA, Soppet DR, Shi Y;

XX WPI: 2001-374441/39.

XX Nucleic acids encoding 24 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating e.g. Gaucher's disease,
 PT Alzheimer's disease, Sjögren's syndrome, Creutzfeldt-Jacob disease,
 PT diabetes mellitus and multiple sclerosis -

PS Disclosure: Page 31-32; 532pp; English.

XX AAD0840478 represent cDNAs corresponding to 24 human secreted
 CC protein genes, and AAE04100-AAE04170 represent the proteins they encode.
 CC AAE04172-AAE04197 represent human secreted protein fragments or variants.
 CC The secreted proteins and their genes are useful for preventing, treating
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.
 CC Pathological conditions can be diagnosed by determining the amount of the
 CC new protein in a sample or by determining the presence of mutations in
 CC the new genes. Specific uses are described for each of the 24 genes,
 CC based on the tissues in which they are most highly expressed, and include
 CC developing products for the diagnosis or treatment of proliferative
 CC disorders, cancer, tumours, foetal and developmental abnormalities,
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
 CC angiotensin disorders, kidney disorders, gastrointestinal disorders,
 CC pregnancy-related disorders, endocrine disorders, and infections. The
 CC proteins can also be used to aid wound healing and epithelial cell
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs
 CC before transplantation, for supporting cell culture of primary tissues,
 CC to regenerate tissues, to identify their cognate ligands or binding
 CC partners, and in chemotaxis, and can be used as a food additive or
 CC preservative to modify storage properties. Antibodies specific for a
 CC protein of the invention can be used in alleviating symptoms associated
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The
 CC present sequence represents a human secreted protein fragment referred
 CC to in the disclosure of the invention.

XX Sequence 367 AA:

Query Match 49.5%; Score 450; DB 22; Length 367;
 Best Local Similarity 50.3%; Pred. No. 3.7e-33;
 Matches 88; Conservative 30; Mismatches 49; Indels 8; Gaps 4;

QY 1 MKDVYKTPQMGDPASLEPQIAETLSNIERLKEVQKAEWLAESAESVLSNRGDSLSRH 60
 DB 193 MKDVYKTPQMGDPASLEPQIAETLSNIERLKEVQKAEWLAESAESVLSNRGDSLSRH 249
 QY 61 ARPPXPAPASPPDSSSSNASQDTERSESEPPSESDPTIYTERDEDEF-EEPTSPGHC 119
 DB 250 ssdlnhltvgrespegsytdangevrpgpqgh---hnefdedddpjpaiqhc 305
 QY 120 VAITHFEGSSRGITSMAGEDLSMEEDKDGWTRVRKKEGEGYVPTSYRLTL 174
 DB 306 Kaiypfdghegtlamkegevllyleedkdgwtrarrqgeegyvplsyldvcl 360

RESULT 8

AA061130 standard; Protein; 547 AA.

XX AAB61130;

XX 30-MAR-2001 (first entry)

XX Human NOV2 protein.

XX

KW Human; NOVX; antiinflammatory; cytoskeletal; neuroprotective;
 KW cerebroprotective; immunomodulator; vulnerary; vasotropic; gene therapy;
 KW hyperplasia; tumour; restenosis; psoriasis; Dupuytren's contracture;
 KW diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease.

OS Homo sapiens.

PN WO200075321-A2.

XX 14-DEC-2000.

XX 01-JUN-2000; 2000WO-US15303.

XX 03-JUN-1999; 99US-0137322.

XX 16-MAR-2000; 2000US-0189810.

XX 22-MAR-2000; 2000US-0191158.

XX 30-MAR-2000; 2000US-0193086.

XX 31-MAY-2000; 2000US-0137322.

XX (CURA-) CURAGEN CORP.

XX Shinketsu RA, Fernandes E, Herrman J, Vernet C;

XX WPI: 2001-102403/11.

XX N-PSDB; AAF27850.

XX New NOVX polypeptides and polynucleotides, useful in gene therapy, as a
 PT diagnostic marker, protein therapeutic, antibody or small molecule drug
 PT target for treating immune, proliferative and metabolic diseases and
 PT wound healing -

XX Claim 1; Page 22-24; 194pp; English.

XX The present sequence is a new isolated polypeptide (NOVX). The NOVX
 CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for
 CC treating or preventing NOVX-associated disorders. They are also useful
 CC for determining the presence of or a predisposition to a disease
 CC associated with altered levels of the NOVX polypeptide or nucleic acid.
 CC These NOVX-associated disorders include hyperplasias, tumours,
 CC restenosis, psoriasis, Dupuytren's contracture, diabetic complications,
 CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral
 CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides
 CC are especially useful in gene therapy. Specifically, NOVX is useful as
 CC a diagnostic marker or prognostic marker, protein therapeutic and
 CC antibody target or small molecule drug target to treat disorders in the
 CC immune response pathway, thyroid and metabolic diseases, bone metabolic
 CC disorders, diseases of the pancreas (e.g. diabetes or digestive
 CC disorders), proliferative diseases, or tissue regeneration and
 CC development (e.g. wound healing or treatment of burns).

XX Sequence 547 AA:

Query Match 49.3%; Score 449; DB 22; Length 547;
 Best Local Similarity 50.3%; Pred. No. 7.6e-33;
 Matches 88; Conservative 30; Mismatches 49; Indels 8; Gaps 4;

QY 1 MKDVYKTPQMGDPASLEPQIAETLSNIERLKEVQKAEWLAESAESVLSNRGDSLSRH 60
 DB 373 MKDVYKTPQMGDPASLEPQIAETLSNIERLKEVQKAEWLAESAESVLSNRGDSLSRH 429
 QY 61 ARPPXPAPASPPDSSSSNASQDTERSESEPPSESDPTIYTERDEDEF-EEPTSPGHC 119
 DB 430 ssdlnhltvgrespegsytdangevrpgpqgh---hnefdedddpjpaiqhc 485
 QY 120 VAITHFEGSSRGITSMAGEDLSMEEDKDGWTRVRKKEGEGYVPTSYRLTL 174
 DB 486 Kaiypfdghegtlamkegevllyleedkdgwtrarrqgeegyvplsyldvcl 540

RESULT 9

AA024922 standard; Protein; 265 AA.

XX

XX AC AAY24922;
 XX DT 27-AUG-1999 (first entry)
 XX DE Human SH3-containing protein 1.
 XX KW Human SH3 containing protein; HS3C-1; HS3C-2; Src homology 3 domain;
 KW diagnosis; cancer; immune disorder; development disorder; leukemia;
 KW immunoflammatory condition; rheumatoid arthritis; ulcerative colitis;
 KW osteoarthritis; Gaucher's disease; adenocarcinoma; lymphoma; melanoma;
 KW sarcoma; AIDS; allergy; asthma; irritable bowel syndrome; pancreatitis;
 KW multiple sclerosis; osteoarthritis; haemodialysis; infection; trauma;
 KW anaemia; epilepsy; congenital glaucoma.
 XX OS Homo sapiens.
 XX PN US916753-A.
 XX PD 29-JUN-1999.
 XX PF 13-NOV-1997; 97US-0970133.
 XX PR 13-NOV-1997; 97US-0970133.
 XX PA (INCY-) INCYTE PHARM INC.
 XX PI Bandman O, Guegler KJ, Lal P;
 XX DR MPI: 1999-394206/33.
 XX DR N-PSDB; AAX83628.
 XX PT New Src homology 3 domain containing proteins useful for the
 XX PT diagnosis, treatment or prevention of cancer and immune or
 XX PT development disorders
 XX PS Claim 1; Fig 1; 32pp; English.
 XX CC The present sequence represents human Src homology 3 domain (SH3)
 CC containing protein 1, designated HS3C-1. HS3C proteins can be used
 CC for the diagnosis, treatment or prevention of cancer and immune or
 CC development disorders. HS3C-1 is particularly expressed in prostate
 CC tissues associated with prostate tumours and HS3C-2 with immunoflammatory
 CC conditions such as rheumatoid arthritis, ulcerative colitis,
 CC osteoarthritis and Gaucher's disease. A vector expressing the complement
 CC of the polynucleotide encoding HS3C-1 can be administered to a subject
 CC to prevent or treat cancers including adenocarcinoma, leukemia, lymphoma,
 CC melanoma, sarcoma and especially cancers of the bladder, kidney, heart,
 CC lung, adrenal gland, skin, spleen, liver, ovary, pancreas, thyroid and
 CC uterus. An immune disorder such as AIDS, allergies, asthma, irritable
 CC bowel syndrome, multiple sclerosis, pancreatitis and osteoarthritis can
 CC also be treated along with complications of cancer, haemodialysis, viral,
 CC bacterial, fungal, and parasitic infections and trauma. A vector
 CC expressing the complement of the polynucleotide encoding HS3C-2 can also
 CC be administered to a subject to prevent or treat cancers and immune
 CC disorders as well as developmental disorders such as anaemia, epilepsy,
 CC and congenital glaucoma. The expression vectors which encode HS3C can be
 CC used to deliver nucleotide sequences to targeted organ, tissue or cell
 CC populations and antisense polynucleotides to treat conditions associated
 CC with overexpression of HS3C by blocking transcription of the mRNA,
 CC modulating HS3C activity or regulating the gene function.
 XX CC
 XX SQ Sequence 265 AA;
 QY Query Match 47.5%; Score 432.5; DB 20; Length 265;
 Db Best Local Similarity 49.5%; Pred. No. 9.8e-32;
 Matches 91; Conservative 28; Mismatches 46; Indels 19; Gaps 4;
 QY 1 MKDYERTPGMGDASLEPQIAFTLSNIERLKLKQKYEAWLAESV-----LSNRGD 55
 Db 82 mkdyvrlknpgmgdpsadlkhkaevsnglekrlvtetqkfeawlaevgrlparnegarrts 141

QY 56 SLSRHARPPXPASAPPDSSNSASQDTKE---SSEPPSESQDPIRYEPEDRF-EE 111
 Db 142 glydsqnp-----lvncaqdrespgsyeegsqesemkyladlddeiddee 191
 QY 112 PTSPIGHCAIYHEGSSSECTISMAGEEDLSLMEEDKGDGWTVRKKEGEGVPTSYLR 171
 Db 192 plpaigtckalytfegqnegtlsvgegtlyvleekdgdwtcrrnedeeglyvpslye 251
 QY 172 VTLN 175
 Db 252 vcld 255
 RESULT 10
 AAB92528
 ID AAB92528 standard; Protein; 537 AA.
 XX AC AAB92528;
 XX DT 26-JUN-2001 (first entry)
 XX DE Human protein sequence SEQ ID NO:10684.
 XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX OS Homo sapiens.
 XX PN EP1074617-A2.
 XX PD 07-FEB-2001.
 XX PF 28-JUL-2000; 2000EP-0116126.
 XX PR 29-JUL-1999; 99JP-0248036.
 XX PR 27-AUG-1999; 99JP-0300253.
 XX PR 11-JAN-2000; 2000JP-0118776.
 XX PR 02-MAY-2000; 2000JP-0183767.
 XX PR 09-JUN-2000; 2000JP-0241899.
 XX PA (HELI-) HELIX RES INST.
 XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX DR MPI: 2001-318749/34.
 XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
 XX PT full-length cDNAs defined in the specification, and for the detection
 XX PT and/or diagnosis of the abnormality of the proteins encoded by the
 XX PT full-length cDNAs -
 XX PS Claim 8; SEQ ID 10684; 2537bp + CD ROM; English.
 XX CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632


```

PR 10-SEP-1998; 98US-0099741.
PR 10-SEP-1998; 98US-0099754.
PR 10-SEP-1998; 98US-0099763.
PR 10-SEP-1998; 98US-0099792.
PR 10-SEP-1998; 98US-0099808.
PR 10-SEP-1998; 98US-0099812.
PR 10-SEP-1998; 98US-0099815.
PR 10-SEP-1998; 98US-0099816.
PR 15-SEP-1998; 98US-0100385.
PR 15-SEP-1998; 98US-0100388.
PR 15-SEP-1998; 98US-0100390.
PR 16-SEP-1998; 98US-0100584.
PR 16-SEP-1998; 98US-0100627.
PR 16-SEP-1998; 98US-0100661.
PR 16-SEP-1998; 98US-0100662.
PR 16-SEP-1998; 98US-0100664.
PR 17-SEP-1998; 98US-0100683.
PR 17-SEP-1998; 98US-0100684.
PR 17-SEP-1998; 98US-0100710.
PR 17-SEP-1998; 98US-0100711.
PR 17-SEP-1998; 98US-0100919.
PR 17-SEP-1998; 98US-0100930.
PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101067.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.

PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0106500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 18-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.

PA (GETH ) GENENTECH INC.
XX Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX WPI; 2000-237871/20.
XX N-PSDB; AAA37110.
XX
XX New mammalian DNA sequences encoding transmembrane, receptor or
XX secreted PRO polypeptides, useful for screening of potential peptide or
XX small molecule inhibitors of the relevant receptor/ligand interactions
XX
XX Claim 12; Fig 178; 773pp; English.
XX
XX AAA37022 to AAA37144 encode the new isolated human transmembrane,
XX receptor or secreted PRO polypeptides given in AA99340 to AA99462. The
XX transmembrane and receptor PRO proteins can be used for screening of
XX potential peptide or small molecule inhibitors of the relevant
XX receptor/ligand interactions. The polypeptides and nucleotide sequences
XX encoding them have various industrial applications, including uses as
XX pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
XX PCR primers and hybridisation probes used in the isolation of the PRO
XX polypeptides from the present invention.
XX
XX Sequence 370 AA;

OY 16 SLEPQIAETLSNIERLKEVCKEYEMIAEESRYSLNQGDLSRRHARPPXPASAPD-- 73
DB 47 sleqlqevreslrraqsvqvkgaarla----llqagldlverwlkpmrlqagdeveqe 101

Query Match 14.5%; Score 131.5; DB 21; Length 370;
Best Local Similarity 26.0%; Pred. No. 0.00049;
Matches 44; Conservative 29; Mismatches 71; Indels 25; Gaps 6;

```


transduction processes, etc. New candidate drugs can be identified, and their specificities (e.g. pharmacological activities) can be assessed using the method of the invention.

Sequence 377 AA;

Query Match 13.6%; Score 123.5; DB 17; Length 377;
Best Local Similarity 23.0%; Pred. No. 0.0027;
Matches 65; Conservative 33; Mismatches 67; Indels 117; Gaps 15;

QY 2 KDVKRTQMDPASPALPOIAETLSNI-----ERLK-----LEVQR----- 37
D 66 kekyskikeldqgt--pqymenmeqvfeeqcqfkeekrtirrfrevllvqkhlslsnvag 153
QY 38 -----YEAWLAESAESVLSNR----- 53
D 154 ykaiyhdiqslraadavedlrfwfranhgpgmammwpqfeesadl-nrlsrrkkst 212
QY 54 -----GDSLSRHARP-----PXPPA-SAPPDSSSN-----SASQDT 83
D 213 dgvlltglnqtdq-slpkpsstlnvpsnpgsqsqsympfedddltgvtsekdtt 271
QY 84 K-----ESSEEPSESSOD---TPI-YTERDEP---FEEPTPIGHCV-AIYHFGS 128
D 272 kakvssyektgypctdwdsenmpfssldangsnpfdtdatsgtevrralydyegq 331
QY 129 SEGTSMAEGEDLSLMEEDKGDGWTVRRRKEGEGYVPTSYL 170
D 332 ehdelstfkagdeltkmededegwckgrlndngqyglypany 373

RESULT 15

AAB43338 standard; Protein: 486 AA.

AC AAB43338;

DT 08-FEB-2001 (first entry)

DE Human ORFX ORF3102 polypeptide sequence SEQ ID NO:6204.

XX Human; open reading frame; ORFX; detection; cytotstatic; hepatotropic;
XX vulnery; antiposrotatic; antiparkinsonian; noctropic; neuroprotective;
XX anticonvulsant; osteoprotic; antiarthritic; immunosuppressant; cardiant;
XX hypotensive; dermatological; immunosuppressive; antidiabetic;
XX antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
XX antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX thrombosis; contraceptive.

XX Homo sapiens.

XX WO200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 07-APR-1999; 99US-0127636.

XX 05-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shinketsu, RA, Leach M;

XX PI

XX WPI: 2000-602362/57.
DR N-PSDB: AAC77547.

XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -

XX Claim 11; Page 5390-5391; 5507pp; English.

XX AAC7446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytotstatic; hepatotropic; vulnery;
XX antiposrotatic; antiparkinsonian; noctropic; neuroprotective;
XX osteoprotic; anticonvulsant; antirheumatic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
XX antithyroid; and antinaemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antiinflammatory disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 486 AA;

Query Match 13.6%; Score 123.5; DB 21; Length 486;
Best Local Similarity 23.0%; Pred. No. 0.0038;
Matches 65; Conservative 33; Mismatches 67; Indels 117; Gaps 15;

QY 2 KDVKRTQMDPASPALPOIAETLSNI-----ERLK-----LEVQR----- 37
D 205 kekyskikeldqgt--pqymenmeqvfeeqcqfkeekrtirrfrevllvqkhlslsnvag 262
QY 38 -----YEAWLAESAESVLSNR----- 53
D 263 ykaiyhdiqslraadavedlrfwfranhgpgmammwpqfeesadl-nrlsrrkkst 321
QY 54 -----GDSLSRHARP-----PXPPA-SAPPDSSSN-----SASQDT 83
D 322 dgvlltglnqtdq-slpkpsstlnvpsnpgsqsqsympfedddltgvtsekdtt 380
QY 84 K-----ESSEEPSESSOD---TPI-YTERDEP---FEEPTPIGHCV-AIYHFGS 128
D 361 kakvssyektgypctdwdsenmpfssldangsnpfdtdatsgtevrralydyegq 440
QY 129 SEGTSMAEGEDLSLMEEDKGDGWTVRRRKEGEGYVPTSYL 170
D 441 ehdelstfkagdeltkmededegwckgrlndngqyglypany 482

Search completed: April 7, 2002, 16:08:16
Job time: 459 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2002, 16:08:42 ; Search time 19.76 Seconds

(without alignments)
199.295 Million cell updates/sec

Title: US-09-925-122A-3

Perfect score: 910
Sequence: 1 MKDYVETPQMGDPASLEPQ.....RRKEGGGYPTSLRVTLN 175

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfillsl.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	908	99.8	175	2	US-08-970-133-3
2	432.5	47.5	265	2	US-08-970-133-1
3	422.5	46.4	237	2	US-08-970-133-5
4	119	13.1	659	1	US-08-391-615-2
5	117	12.9	659	5	PCT-US95-05008-8
6	114.5	12.6	739	4	US-09-022-983-2
7	108	11.9	533	1	US-07-820-011A-2
8	108	11.9	533	5	PCT-US93-00445-2
9	106	11.6	442	1	US-08-391-615-6
10	105	11.5	620	5	PCT-US95-05008-9
11	104.5	11.5	496	2	US-09-006-675-2
12	104.5	11.5	496	4	US-09-228-603A-2
13	104	11.4	536	1	US-07-820-011A-4
14	104	11.4	536	5	PCT-US93-00445-4
15	104	11.4	536	5	PCT-US95-05008-13
16	102.5	11.3	450	5	PCT-US95-05008-12
17	101.5	11.2	450	5	PCT-US95-05008-7
18	100	11.0	821	1	US-07-935-111A-4
19	100	11.0	821	1	US-08-368-079-4
20	100	11.0	821	5	PCT-US93-0796-4
21	99.5	10.9	543	2	PCT-US95-05008-14
22	98.5	10.8	546	2	US-08-942-423-4
23	96.5	10.6	401	2	US-08-549-004A-5
24	96.5	10.6	401	4	US-09-051-982A-5
25	95	10.4	60	1	US-08-627-197-1
26	94.5	10.3	400	1	US-08-938-830-29
27	94	10.3	537	5	PCT-US95-05008-11

28	93	10.2	625	1	US-08-391-615-3	Sequence 3, Appl1
29	91	10.0	529	5	PCT-US95-05008-15	Sequence 15, Appl1
30	90	9.9	740	4	US-09-022-983-5	Sequence 5, Appl1
31	89	9.8	480	3	US-09-189-035-5	Sequence 5, Appl1
32	89	9.8	480	4	US-09-382-086-5	Sequence 4, Appl1
33	88.5	9.7	228	1	US-08-167-035-47	Sequence 47, Appl1
34	88.5	9.7	228	1	US-08-208-887A-47	Sequence 47, Appl1
35	88.5	9.7	228	1	US-08-539-005-47	Sequence 47, Appl1
36	88.5	9.7	228	2	US-08-815-176-5	Sequence 5, Appl1
37	88.5	9.7	486	2	US-08-942-423-3	Sequence 3, Appl1
38	88.5	9.7	493	2	US-08-933-821-2	Sequence 2, Appl1
39	88.5	9.7	493	3	US-08-960-507-2	Sequence 2, Appl1
40	88.5	9.7	512	5	PCT-US95-05008-16	Sequence 16, Appl1
41	88	9.7	49	2	US-08-459-568-53	Sequence 53, Appl1
42	88	9.7	49	2	US-08-399-411-53	Sequence 53, Appl1
43	88	9.7	49	3	US-08-516-859A-53	Sequence 53, Appl1
44	87	9.6	261	2	US-08-691-814B-4	Sequence 4, Appl1
45	86.5	9.5	54	1	US-08-167-035-28	Sequence 28, Appl1

ALIGNMENTS

RESULT 1
US-08-970-133-3
; Sequence 3, Application US/08970133
; Patent No. 5916753
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Guegler, Karl J.
; TITLE OF INVENTION: SH3-CONTAINING PROTEINS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/970,133
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0419 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 175 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSNOT20
; CLONE: 1816529
; US-08-970-133-3

Query Match 99.8% Score 908; DB 2; Length 175;
Best Local Similarity 100.0%; Pred. No. 2.8e-81;
Matches 175; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MKDYYEKIPQMGDASLEPQIAETLSNTERKLEVOYKTEAMLAEBERYLSNKGDSLSRH	60
Db	1	MKDYYEKIPQMGDAPASLEPQIAETLSNTERKLEVOYKTEAMLAEBERYLSNKGDSLSRH	60
QY	61	ARPPXPASAPDDSSNSASODTKESSEEPSPSESOPTPYTDEDDDFEEBPTSPIGHCV	120
Db	61	ARPPXPASAPDDSSNSASODTKESSEEPSPSESOPTPYTDEDDDFEEBPTSPIGHCV	120
QY	121	AIYHFGESSECTISMAEGEDLSLNEEDKGDQMTVRKKEGEGVVPSTYLRVTLN	175
Db	121	AIYHFGESSECTISMAEGEDLSLNEEDKGDQMTVRKKEGEGVVPSTYLRVTLN	175

```

1      RESULT 2
2      US-08-970-133-1
3      : Sequence 1, Application US/08970133
4      : Patent No. 5916753
5      :
6      : GENERAL INFORMATION:
7      : APPLICANT: Bandman, Olga
8      : APPLICANT: Gueglar, Karl J.
9      : APPLICANT: Lal, Preeti
10     : TITLE OF INVENTION: SH3-CONTAINING PROTEINS
11     : NUMBER OF SEQUENCES: 5
12     : CORRESPONDENCE ADDRESS:
13     : ADDRESSEE: Incyte Pharmaceuticals, Inc.
14     : STREET: 3174 Porter Dr.
15     : CITY: Palo Alto
16     : STATE: CA
17     : COUNTRY: USA
18     : ZIP: 94304
19     :
20     : COMPUTER READABLE FORM:
21     : MEDIUM TYPE: Diskette
22     : COMPUTER: IBM Compatible
23     : OPERATING SYSTEM: DOS
24     : SOFTWARE: FastSeq for Windows Version 2.0
25     :
26     : CURRENT APPLICATION DATA:
27     : APPLICATION NUMBER: US/08/970,133
28     : FILING DATE: Filled Herewith
29     : PRIOR APPLICATION DATA:
30     : APPLICATION NUMBER:
31     : FILING DATE:
32     :
33     : ATTORNEY/AGENT INFORMATION:
34     : NAME: Billings, Lucy J.
35     : REGISTRATION NUMBER: 36,749
36     : REFERENCE/DOCKET NUMBER: PF-0419 US
37     : TELECOMMUNICATION INFORMATION:
38     : TELEPHONE: 650-855-0555
39     : TELEFAX: 650-845-4166
40     :
41     : INFORMATION FOR SEQ. ID NO: 1:
42     : SEQUENCE CHARACTERISTICS:
43     : LENGTH: 265 amino acids
44     : TYPE: amino acid
45     : STRANDEDNESS: single
46     : TOPOLOGY: linear
47     : IMMEDIATE SOURCE:
48     : LIBRARY: BRAITVUT03
49     : CLONE: 865744
50     :
51     : US-08-970-133-1

```

[illegible][illegible]

```

? RESULT 3
? US-08-970-133-5
? Sequence 5, Application US/08970133
? Patent No. 5916753
? GENERAL INFORMATION:
? APPLICANT: Bandman, Olga
? APPLICANT: Aueglar, Karl J.
? APPLICANT: Lal, Preeti
? TITLE OF INVENTION: SH3-CONTAINING PROTEINS
? NUMBER OF SEQUENCES: 5
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Incyte Pharmaceuticals, Inc.
? STREET: 3174 Porter Dr.
? CITY: Palo Alto
? STATE: CA
? COUNTRY: USA
? ZIP: 94304
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FASTSEQ for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08-970,133
? FILING DATE: Filed Herewith
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Billings, Lucy J.
? REGISTRATION NUMBER: 36,749
? REFERENCE/DOCKET NUMBER: PF-0419 US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 650-855-0555
? TELEFAX: 650-845-4166
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 237 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? IMMEDIATE SOURCE:
? LIBRARY: GenBank
? CLONE: 1255033
? US-08-970-133-5

```

	Query Match	46.4%	Score 422.5;	DB 2;	Length 237;
	Best Local Similarity	49.2%;	Pred. No. 9.6e-34;		
	Matches 86;	Conservative 27;	Mismatches 35;	Indels 9;	Gaps 3.
OY	1 MKDYVEKTPQMGDADSLPEQIAETLSNIERLKEVQKYEAMIAEASRYLSNRGDSLRH	60			
	: :				
Dd	54 MKDYVLKNPQMGDPDSLQDKLEVTYQNIEKLRLKQKFEAMIAEVEGRLL-----PARSRQ	108			
OY	61 ARPPXPPASAPDDSSSNSASODTKE---SSEBPSESODPIIYEFEDF--EEEPSTSI	116			
	: :				
Dd	109 ARROSGLYDGTHOTIVNCAODRSPDGSYTUEEGSQESEHNKYLADFDFDEFDEBPLPAI	168			
OY	117 GHCAVAIHFEBSRGTTISMAEGDLISLMEDKKGDGWTVRRKRGEAGYPTSYLKATLNL	175			
	: : :				
Dd	169 GTCCALVTEESQNGTSIVSEGTTISYLEEKKGDDWTVIRRNDEGGYSYVVEUYLD	227			


```

RESULT      4
US-08-391-615-2
; Sequence 2, Application US/08391615
; Patent No. 5550054
; GENERAL INFORMATION:
;   APPLICANT: Witte, Owen
;   APPLICANT: Tsukada, Satoshi
;   APPLICANT: Safiran, Douglas
;   APPLICANT: Rawlings, David
;   TITLE OF INVENTION: HEMATOPOIETIC RESTRICTED TYROSINE KINASE
;   NUMBER OF SEQUENCES: 7
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
;     STREET: 4 Embarcadero Center, Suite 3400
;     CITY: San Francisco
;     STATE: California
;     COUNTRY: USA
;     ZIP: 94111-4187
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/391,615
;     FILING DATE:
;     CLASSIFICATION: 435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: US 08/006,449
;       FILING DATE: 21-JAN-1993
;       ATTORNEY/AGENT INFORMATION:
;         NAME: Rowland, Bertam I
;         REGISTRATION NUMBER: 20,015
;         REFERENCE/DOCKET NUMBER: A-57583-1/BIR UCLA 232-1
;         TELECOMMUNICATION INFORMATION:
;           TELEPHONE: (415) 781-1989
;           TELEFAX: (415) 398-3249
;           TELEX: 910 277299 FHT UR
;       INFORMATION FOR SEQ ID NO: 2:
;         SEQUENCE CHARACTERISTICS:
;           LENGTH: 659 amino acids
;           TYPE: amino acid
;           TOPOLOGY: linear
;         MOLECULE TYPE: protein
; US-08-391-615-2

Query Match      13.1%; Score 119; DB 1; Length 659;
Best Local Similarity 23.3%; Pred. No. 0.0015;
Matches 44; Conservative 31; Mismatches 68; Indels 46; Gaps 9;

QY      4 VYEKRP-QMGDPASLEPOIAETLSNIEELKE-VQKYEA-WL-----AAES--- 47
      ||:|:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      105 VIDSEPLLYVFSTELERKRWIHQLKNVIRYNSDLVQKTHPCWIDGQYICSCQTKNMAG 164
      - - - - - RHARPPXPASAPDPSSSNSASQDKESSESEPPSESDTPY 101
QY      48 -RVLSNRGDSLS-----RHARPPXPASAPDPSSSNSASQDKESSESEPPSESDTPY 101
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      165 CQILENRNGSLKPGSSHKRTKKPLP-----TPREDQLKKPLPEPTAAPI- 211
      - - - - - STTELKKVVALDYDMPMANADQLQKRGEEYFILESSNLP-WWRARDKNQO 260
QY      102 TEFDEDFEPEPSPIGHCAVAIYHEFGSSSEGTISMAEGEDLSIMEDKDGQWTRVRKKEG 161
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      212 -----STTELKKVVALDYDMPMANADQLQKRGEEYFILESSNLP-WWRARDKNQO 260
      : : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      162 EGYVPTSYL 170
      ||:|:|:|:|:|
DB      261 EGYIPSNVI 269
      ||:|:|:|:|:|

RESULT      5
PCT-US95-05008-8
; Sequence 8, Application PC/TUS9505008
; GENERAL INFORMATION:

```

[illegible]

```
Patent No. 6159731
GENERAL INFORMATION:
APPLICANT: Yang, Xiaolu
APPLICANT: Khosravi-Far, Roya
APPLICANT: Chang, Howard Y.
APPLICANT: Baltimore, David
TITLE OF INVENTION: DAXX, A NOVEL FAS-BINDING
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESS: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022,983
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/037,919
FILING DATE: 12-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/051,753
FILING DATE: 26-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Van Amsterdam, John R.
REGISTRATION NUMBER: 40,212
REFERENCE/DOCKET NUMBER: M0656/7036
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 739 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-09-022-983-2

Query Match      12.6%; Score 114.5; DB 4; Length 739;
Best Local Similarity 27.2%; Pred. No. 0.0049;
Matches 47; Conservative 20; Mismatches 67; Indels 39; Gaps 7;

QY 13 DPASLEPQIAETLSN-----IERKLEVOKYEMLAELASRVLSNCGDSLRIHAR--PPX 65
DB 355 DPLSLPTLARRLRRENTLARNRLDLVISKYAMODKTE-----EGERQRRARLLGTA 408
QY 66 PPAASAPPDSSNS-----ASODTKESSEEPSESDPTPIYTEDEDFEEEPPTSPIGH 118
DB 409 PPSPPPPQASSSGEGPSMASOECTTKAETDDDDDDDDDDDDDDDDDDDDDDDDDDDDDD 463
QY 119 CVAIHFEGSSGEGTSMAGEDLSLMEEDKGDGWTFRKKEGE---GYVPTS 168
DB 464 -----EEEEKKATDEDEDELEQLQEDQGD---EEEGGDNGNSPTS 505

RESULT 7
US-07-820-011A-2
Sequence 2, Application US/07820011A
Patent No. 5336615
GENERAL INFORMATION:
APPLICANT: Bell, Leonard
APPLICANT: Madril, Joseph A.
```

```
APPLICANT: Warren, Stephen L.
APPLICANT: Luthringer, Daniel J.
TITLE OF INVENTION: Genetically Engineered
TITLE OF INVENTION: Endothelial Cells Exhibiting Enhanced
TITLE OF INVENTION: Migration
TITLE OF INVENTION: and Plasminogen Activator Activity
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Maurice M. Klee
STREET: 1951 Burr Street
CITY: Fairfield
STATE: Connecticut
COUNTRY: USA
ZIP: 06430
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 kb storage
COMPUTER: IBM PC XT
OPERATING SYSTEM: PC-DOS/MS-DOS 2.10
SOFTWARE: Displaywrite 3
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/820,011A
FILING DATE: 19920106
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Klee, Maurice M.
REGISTRATION NUMBER: 30,399
REFERENCE/DOCKET NUMBER: LB-101
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203) 254 1101
TELEFAX: (203) 255 1400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 533 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: Protein
HYPOTHETICAL: No
FRAGMENT TYPE: Complete Sequence
ORIGINAL SOURCE:
ORGANISM: Gallus, gallus
PUBLICATION INFORMATION:
AUTHORS: Takeya, Tatsuo
AUTHORS: Hanafusa, Hidesaburo
TITLE: Structure and Sequence of the
TITLE: Cellular Gene Homologous to the RSV src
TITLE: Gene and the Mechanism for Generating the
TITLE: Transforming Virus
JOURNAL: Cell
VOLUME: 32
PAGES: 881-890
DATE: March, 1983
US-07-820-011A-2

Query Match      11.9%; Score 108; DB 1; Length 533;
Best Local Similarity 22.6%; Pred. No. 0.014;
Matches 31; Conservative 27; Mismatches 53; Indels 26; Gaps 4;

QY 58 SRHARPPXP-----ASAPPDS-----SNSASQDTKESSEEPSESDPTPIYT 102
DB 3 SSKSKPKRPSQRRSLRLEPDSITHGGFPASQTPNKTAAPDTHRTPSRFGVATEPKLIG 62
QY 103 EPEDEEEETSP-----IGHCVAIHFEGSSGEGTSMAGEDLSLMEEDKGDGWT 153
DB 63 GFNT--SDTVTSPQRAGALAGVTFVALYDYSRTETDLSFKKGRLOIVNTEGDWML 120
QY 154 RVRKKEGEGVYPTSYL 170
DB 121 AHSLTGTGQYIPSNV 137

RESULT 8
PCT-US93-00445-2
```

[illegible]

```

1 RESULT 10
2 PCT-US95-05008-9
3 ; Sequence 9, Application PC/TUS9505008
4 ; GENERAL INFORMATION:
5 ; APPLICANT: Sugen, Inc.
6 ; APPLICANT: 515 Galveston Drive
7 ; APPLICANT: Redwood City, California 94063-4720
8 ; APPLICANT: United States of America
9 ; APPLICANT: Wissenschaften E.V.
10 ; APPLICANT: Hoigarten Str. 2
11 ; APPLICANT: Munchen 80539
12 ; APPLICANT: Germany
13 ; TITLE OF INVENTION: Novel Megakaryocytic Protein Tyrosine
14 ; TITLE OF INVENTION: Kinases
15 ; NUMBER OF SEQUENCES: 21
16 ; CORRESPONDENCE ADDRESS:
17 ; ADDRESSEE: Pennie & Edmonds
18 ; STREET: 1155 Avenue of the Americas
19 ; CITY: New York
20 ; STATE: New York
21 ; COUNTRY: U.S.A.
22 ; ZIP: 10036
23 ;
24 ; COMPUTER READABLE FORM:
25 ; MEDIUM TYPE: Floppy disk
26 ; COMPUTER: IBM PC compatible
27 ; OPERATING SYSTEM: PC-DOS/MS-DOS
28 ; SOFTWARE: Patentin Release #1.0, Version #1.25
29 ; CURRENT APPLICATION DATA:
30 ; APPLICATION NUMBER: PCT/US95/05008
31 ; FILING DATE: 24-APR-1995
32 ;
33 ; CLASSIFICATION:
34 ;
35 ; PRIOR APPLICATION DATA:
36 ; APPLICATION NUMBER: US 08/232,545
37 ; FILING DATE: 22-APR-1994
38 ; CLASSIFICATION:
39 ;
40 ; ATTORNEY/AGENT INFORMATION:
41 ; NAME: Coruzzi, Laura A.
42 ; REGISTRATION NUMBER: 30,742
43 ; REFERENCE/DOCKET NUMBER: 7683-074
44 ; TELECOMMUNICATION INFORMATION:
45 ; TELEPHONE: (212)790-9090
46 ; TELEFAX: (212)869-9741
47 ;
48 ; TELEX: 66141 PENNIE
49 ;
50 ; INFORMATION FOR SEQ ID NO: 9:
51 ; SEQUENCE CHARACTERISTICS:
52 ; LENGTH: 620 amino acids
53 ; TYPE: amino acid
54 ; STRANDEDNESS: unknown
55 ; TOPOLOGY: unknown
56 ;
57 ; MOLECULE TYPE: protein
58 ;
59 PCT-US95-05008-9

```

```

Query Match      11.5%; Score 105; DB 5; Length 620;
Best Local Similarity 27.8%; Pred No. 0.033;
Matches 25; Conservative 22; Mismatches 29; Indels 14; Gaps 4;

QY      83   TRKSEEE--PSEESODPIYTEFDEDEEPTSPICGVAIYHFGSSEGTISMAEGED 140
          ||:::|||:-||:::||::|||:-||::||:-||::||:-||::||:-||:
Db       149  TKNSKKRLPPTPEDNRRPLW-----EEELV--VIALIDYTNDPELALRNDE 197

QY      141  LSLMEEDKGDCMTRYRNRGEGGGYPTSYL 170
          :::::-||:-||:-||::|||::|||
Db       198  YCLDSSSE-IHMWRVQDRNGHEGYVPSSYL 226

RESULT 11
US-09-006-675-2
; Sequence 2, Application US/09006675
; Patent No. 5952213
; GENERAL INFORMATION:
; APPLICANT: Hemmati-Briavanlou, Ali
; APPLICANT: Weinstein, Daniel C.
```

```

1  TITLE OF INVENTION:  A NOVEL SRC-FAMILY KINASE AND METHODS OF
2  TITLE OF INVENTION:  USE THEREOF
3  NUMBER OF SEQUENCES:  12
4  CORRESPONDENCE ADDRESS:
5  ADDRESSEE:  Klauber & Jackson
6  STREET:  411 Hackensack Avenue, 4th Floor
7  CITY:  Hackensack
8  STATE:  New Jersey
9  COUNTRY:  USA
10 ZIP:  07601
11 COMPUTER READABLE FORM:
12 MEDIUM TYPE:  Floppy disk
13 COMPUTER:  IBM PC Compatible
14 OPERATING SYSTEM:  PC-DOS/MS-DOS
15 SOFTWARE:  PatentIn Release #1.0, Version #1.30
16 CURRENT APPLICATION DATA:
17 APPLICATION NUMBER:  US/09/006,675
18 FILING DATE:  13-JAN-1998
19 CLASSIFICATION:
20 ATTORNEY/AGENT INFORMATION:
21 NAME:  Jackson Esq., David A.
22 REGISTRATION NUMBER:  26,742
23 REFERENCE/DOCKET NUMBER:  600-1-217
24 TELECOMMUNICATION INFORMATION:
25 TELEPHONE:  201-487-5800
26 TELEFAX:  201-343-1684
27 TELEX:  133521
28 INFORMATION FOR SEQ ID NO:  2:
29 SEQUENCE CHARACTERISTICS:
30 LENGTH:  496 amino acids
31 TYPE:  amino acid
32 TOPOLOGY:  linear
33 MOLECULE TYPE:  protein
34 IS-09-006-675-2

```

```
Query Match                                11.5%; Score 104.5; DB 2; Length 496;
Best Local Similarity                      25.0%; Pred. No. 0.027;
Matches   27; Conservative                 23; Mismatches    31; Indels   27; Gaps     4.

QY      71 PPSSSSNSASOD-----TK-ESSEPPPEESSOPTPIYTEFDEDFEEPTSPICGVAI 122
       ||::| : |          || | | : | | | : |
Db      18 PPSQTGLHYVADPSTVTMTKPKRSKRPREGGEVV-----LLAL 59
QY      123 YHFGSSEGTSMAEGEDLSLMEEDKGDGWTFVRRKKEGGEGVPTSYL 170
       ||::| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      60 YDYDGVHPGDLTFRKGDHL-LLKRESGEWMEXELCTISGTGEGFPVSNY 106

RESULT 12
US-09-228-603A-2
; Sequence 2, Application US/09228603A
; Patent No. 6291651
; GENERAL INFORMATION:
APPLICANT: Hemmati-Bilvanlou, Ali
APPlicant: Weinstein, Daniel C.
TITLE OF INVENTION: A NOVEL SRC-FAMILY KINASE AND METHODS OF
TITLE OF INVENTION: USE THEREOF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDressee: Klauder & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
City: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/228,603A
```

FILING DATE: 12-JAN-1999
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-217 N
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-228-603A-2

Query Match 11.5%; Score 104.5; DB 4; Length 496;
Best Local Similarity 25.0%; Pred. No. 0.027;
Matches 27; Conservative 23; Mismatches 31; Indels 27; Gaps 4;

QY 71 PPSSNSASQD-----TK-ESSEPPSESQDTPITTEFDEDEFEETPTSGHCAI 122
DB 18 PPSSTQTHVKNQPTSTVTMTKPKERSKHPREBQGEV-----LLAL 59

QY 123 YHESSECTISMAEGEDLSMEEDKGDGWTVRKKEGEGVPTSYL 170
DB 60 YYYGVHPEDLTKRKDHL-LTKESGEWMEACLSTGEGFVPSNYV 106

RESULT 13
US-07-820-011A-4

; Sequence 4, Application US/07820011A
; Patent No. 5336615

; GENERAL INFORMATION:
; APPLICANT: Bell, Leonard

; APPLICANT: Madril, Joseph A.
; APPLICANT: Warren, Stephen L.

; APPLICANT: Luthringer, Daniel J.
; TITLE OF INVENTION: Genetically Engineered

; TITLE OF INVENTION: Endothelial Cells Exhibiting Enhanced
; TITLE OF INVENTION: Migration

; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Maurice M. Klee
; STREET: 1951 Burr Street

; CITY: Fairfield
; STATE: Connecticut

; COUNTRY: USA
; ZIP: 06430

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb storage

; COMPUTER: IBM PC XT
; OPERATING SYSTEM: PC-DOS/MS-DOS 2.10
; SOFTWARE: Displaywrite 3

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/820,011A

; FILING DATE: 19920106
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:
; NAME: Klee, Maurice M.

; REGISTRATION NUMBER: 30,399
; REFERENCE/DOCKET NUMBER: LB-101

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203) 255 1400
; TELEFAX: (203) 254 1101

; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:

; LENGTH: 536 amino acids
; TYPE: AMINO ACID

; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; HYPOTHETICAL: No
; FRAGMENT TYPE: Complete Sequence
; ORIGINAL SOURCE:

; ORGANISM: Homo sapien
; PUBLICATION INFORMATION:

; AUTHORS: Anderson, Stephen K.
; AUTHORS: Gibbs, Carol P.

; AUTHORS: Tanaka, Akio
; AUTHORS: Kung, Hsing-Jien

; TITLE: Human Cellular src Gene:
; TITLE: Nucleotide Sequence and Derived Amino

; TITLE: Acid Sequence of the Region Coding for
; TITLE: the Carboxy-Terminal Two-Thirds of

; TITLE: pp60c-src
; JOURNAL: Molecular and Cellular Biology

; VOLUME: 5
; ISSUE: 5

; PAGES: 1122-1129
; DATE: May, 1985

; PUBLICATION INFORMATION:
; AUTHORS: Tanaka, Akio

; AUTHORS: Gibbs, Carol P.
; AUTHORS: Arthur, Richard R.

; AUTHORS: Anderson, Stephen K.
; AUTHORS: Kung, Hsing-Jien

; TITLE: DNA Sequence Encoding the
; TITLE: Amino-Terminal Region of the Human c-src

; TITLE: Protein: Implications of Sequence
; TITLE: Divergence among src-Type Kinase

; JOURNAL: Molecular and Cellular Biology
; VOLUME: 7

; ISSUE: 5
; PAGES: 1978-1983

; DATE: May, 1987
; US-07-820-011A-4

Query Match 11.4%; Score 104; DB 1; Length 536;
Best Local Similarity 25.7%; Pred. No. 0.034;
Matches 29; Conservative 20; Mismatches 50; Indels 14; Gaps 3;

QY 67 PASAPDSSNSASQDTKESSEPPSEESQDTPITTEFDEDEFEETPTSP-----IG 117
DB 33 PASQTP---SKPASADGHRGSAAPAAAPKLFGGFNS--SDTVSPORAGPLAGGYT 87

QY 118 HCVAIYHFESSSGTISMAEGEDLSMEEDKGDGWTVRKKEGEGVPTSYL 170
DB 88 TFVALDYESTRTEDLSFKKGERLQIVNTEGDMWLAAHSISTGTGTIPSNYV 140

RESULT 14
PCR-US93-00445-4

; Sequence 4, Application PC/TUS9300445
; GENERAL INFORMATION:

; APPLICANT: Bell, Leonard
; APPLICANT: Madril, Joseph A.

; APPLICANT: Warren, Stephen L.
; APPLICANT: Luthringer, Daniel J.

; TITLE OF INVENTION: Genetically Engineered
; TITLE OF INVENTION: Endothelial Cells

; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Maurice M. Klee
; STREET: 1951 Burr Street

; CITY: Fairfield
; STATE: Connecticut

; COUNTRY: USA
; ZIP: 06430

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 760 Kb storage
COMPUTER: DELL 486/50
OPERATING SYSTEM: DOS 5.0
SOFTWARE: Displaywrite 3
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/00445
FILING DATE: 19930105
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/820,011
FILING DATE: 06-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Klee, Maurice M.
REGISTRATION NUMBER: 30,399
REFERENCE/DOCKET NUMBER: ALX-101PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203) 255 1400
TELEFAX: (203) 254 1101
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 536 amino acids
TYPE: AMINO ACID
TOPOLOGY: Linear
MOLECULE TYPE: Protein
HYPOTHETICAL: NO
FRAGMENT TYPE: Complete Sequence
ORIGINAL SOURCE:
ORGANISM: Homo sapien
PUBLICATION INFORMATION:
AUTHORS: Anderson, Stephen K.
AUTHORS: Gibbs, Carol P.
AUTHORS: Tanaka, Akio
AUTHORS: Kung, Hsing-Jien
AUTHORS: Fujita, Donald J.
TITLE: Human Cellular src gene:
TITLE: Nucleotide Sequence and Derived Amino
TITLE: Acid Sequence of the Region Coding for
TITLE: the Carboxy-Terminal Two-Thirds of
TITLE: pp60c-src
JOURNAL: Molecular and Cellular Biology
VOLUME: 5
ISSUE: 5
PAGES: 1122-1129
DATE: May, 1985
PUBLICATION INFORMATION:
AUTHORS: Tanaka, Akio
AUTHORS: Gibbs, Carol P.
AUTHORS: Arthur, Richard R.
AUTHORS: Anderson, Stephen K.
AUTHORS: Kung, Hsing-Jien
AUTHORS: Fujita, Donald J.
TITLE: DNA Sequence Encoding the
TITLE: Amino-Terminal Region of the Human c-src
TITLE: Protein: Implications for Sequence
TITLE: Divergence among src-Type Kinase
TITLE: Oncogenes
JOURNAL: Molecular and Cellular Biology
VOLUME: 7
ISSUE: 5
PAGES: 1978-1983
DATE: May, 1987
PCT-US93-00445-4

Query Match 11.4%; Score 104; DB 5; Length 536;
Best Local Similarity 25.7%; Pred. No. 0.034;
Matches 29; Conservative 20; Mismatches 50; Indels 14; Gaps 3;

QY 67 PASAPDSSNSASODTKSSSEPPSEESQDPIYTFEDFEEPTSP-----IG 117
DB 33 PASQTP---SKPASADGHGSPSAFAFAAEPKLFGEFNS--SDVTSPQRAQPLAGVT 87

QY 118 HCVAIYHFEQSGESGITISMAEGEDLSLMEEDKGDGWTVRRRKGGEGYVPTSYL 170
DB 88 TEVALYDYESRTEETDLSFKKGERLQIVNTEGDMWLAHSLSTGOTGYIPSNV 140

RESULT 15
PCT-US95-05008-13
Sequence 13, Application PC/TUS9505008
GENERAL INFORMATION:
APPLICANT: Sugen, Inc.
APPLICANT: 515 Galveston Drive
APPLICANT: Redwood City, California 94063-4720
APPLICANT: United States of America
APPLICANT: Wissenschaften E.V.
APPLICANT: Hofgarten Str. 2
APPLICANT: Munchen 80539
APPLICANT: Germany
TITLE OR INVENTION: Novel Megakaryocytic Protein Tyrosine
TITLE OF INVENTION: Kinases
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05008
FILING DATE: 24-APR-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/232,545
FILING DATE: 22-APR-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Cortuzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7683-074
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)790-9090
TELEFAX: (212)869-9741
TELEX: 66141 PENNITE
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 536 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: Protein
PCT-US95-05008-13

Query Match 11.4%; Score 104; DB 5; Length 536;
Best Local Similarity 25.7%; Pred. No. 0.034;
Matches 29; Conservative 20; Mismatches 50; Indels 14; Gaps 3;

QY 67 PASAPDSSNSASODTKSSSEPPSEESQDPIYTFEDFEEPTSP-----IG 117
DB 33 PASQTP---SKPASADGHGSPSAFAFAAEPKLFGEFNS--SDVTSPQRAQPLAGVT 87

QY 118 HCVAIYHFEQSGESGITISMAEGEDLSLMEEDKGDGWTVRRRKGGEGYVPTSYL 170
DB 88 TEVALYDYESRTEETDLSFKKGERLQIVNTEGDMWLAHSLSTGOTGYIPSNV 140

Mon Apr 8 06:09:16 2002

Job time: 171 sec

us-09-925-122a-3.ra1

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2002, 16:09:14 ; Search time 24.96 Seconds

(without alignments)
534.076 Million cell updates/sec

Title: US-09-925-122a-3

Perfect score: 910

Sequence: 1 MDVYEKTPQMGDPASLEPQ.....RRKEGEGVPTSLRYTLN 175

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	447.5	49.2	496	2	JC5261
2	422.5	46.4	237	2	S64718
3	221	24.3	554	2	T15992
4	177.5	19.5	783	2	T23452
5	169.5	18.6	785	2	T23456
6	156.5	17.2	603	2	T43489
7	152	16.7	642	2	T39376
8	130	14.3	633	2	S48956
9	119	13.1	659	2	T49553
10	118	13.0	731	2	T08855
11	117	12.9	659	2	A45184
12	116	12.7	532	1	A34104
13	113.5	12.5	532	1	B34104
14	112.5	12.4	1219	2	S45470
15	112	12.3	1217	2	T39427
16	111	12.2	631	2	T56987
17	109.5	12.0	1119	2	T50995
18	108.5	11.9	498	2	H84484
19	108.5	11.9	1111	1	A33284
20	108	11.9	533	1	TVCHS
21	108	11.9	557	1	TVCFS2
22	108	11.9	568	1	TVCFS1
23	108	11.9	587	1	TVCFSR
24	107	11.8	539	1	T42993
25	107	11.8	4063	2	T23630
26	106.5	11.7	4101	2	TVCFSR
27	106.5	11.6	526	1	OKFVR
28	106	11.6	526	1	OKFVR
29	106	11.6	526	2	S15382

30	106	11.6	526	2	S20808	protein-tyrosine k
31	105.5	11.6	1107	1	S52517	myosin I heavy cha
32	105	11.5	137	2	D24773	protein-tyrosine k
33	105	11.5	620	1	S33253	protein-tyrosine k
34	104.5	11.5	443	2	T27877	hypothetical prote
35	104.5	11.5	545	2	S52313	hypothetical prote
36	104.5	11.5	630	1	T01380	protein-tyrosine k
37	104	11.4	541	2	S31645	protein-tyrosine k
38	103.5	11.4	1003	2	T19533	hypothetical prote
39	103.5	11.4	1123	2	A38962	kinase-related tra
40	103.5	11.4	1130	1	TVHDA	protein-tyrosine k
41	103	11.3	290	2	T42526	hypothetical prote
42	103	11.3	537	1	A43806	protein-tyrosine k
43	102.5	11.3	450	2	A41973	protein-tyrosine k
44	102.5	11.3	536	2	S33569	protein-tyrosine k
45	102.5	11.3	1146	2	B35962	protein-tyrosine k

ALIGNMENTS

RESULT 1
JC5261
salt-tolerant protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 25-Apr-1997 #sequence_revision 09-May-1997 #text_change 09-May-1997
C:Accession: JC5261
R:Tsujii, E.; Tsujii, Y.; Misumi, Y.; Fujita, A.; Sasaguri, M.; Ideishi, M.; Arakawa, K
Biochem. Biophys. Res. Commun. 229, 134-138, 1996
A:Title: Molecular cloning of a novel rat salt-tolerant protein by functional complem
A:Reference number: JC5261; MUID:97112415
A:Accession: JC5261
A:Molecule type: DNA
A:Residues: 1-496 <TSU>
A:Cross-references: DBJ:U50557
A>Note: the authors translated the codon GCA for residue 155 as Arg
A:Comment: This protein is involved in a hormone-dependent transcriptional pathway.
A:Genetics:
A:Gene: STP

Query Match 49.2% Score 447.5; DB 2; Length 496;
Best Local Similarity 83.7% Pred. No. 6.3e-27;
Matches 87; Conservative 5; Mismatches 9; Indels 3; Gaps 1;

OY 1 MDVYEKTPQMGDPASLEPQIAFTLSNIRLKLKYQYKAMLAELASRLVLSNRGSLSRH 60
|||||
Db 371 MDVYEKTPQMGDPASLEPQIAFTLSNIRLKLKYQYKAMLAELASRLVLSNRGSLSRH 430
OY 61 APPXPAPASAPDSSS---NSASQDKESSSEPPSPESQDTPY 101
||| |||
Db 431 TRPPDPPTTAAPPDSSSSNNSSGQDNMESSEPPSEEGDTPY 474
||| |||

RESULT 2
S64718
formin-binding protein 17 - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 05-Nov-1999
C:Accession: S64718; S64711
R:Chan, D.C.; Bedford, M.T.; Leder, P.
submitted to the EMBL Data Library, November 1995
A:Reference number: S64718
A:Accession: S64718
A:Molecule type: mRNA
A:Residues: 1-237 <CHA>
A:Cross-references: EMBL:U40751; NID:q1255032; PIDN:AAC52479.1; PID:q1255033
R:Chan, D.C.; Bedford, M.T.; Leder, P.
EMBO J. 15, 1045-1054, 1996
A:Title: Formin binding proteins bear WW/PW domains that bind proline-rich peptides
A:Reference number: S64711; MUID:96183189
A:Accession: S64711
A>Status: nucleic acid sequence not shown

F:2-131/Domain: pleckstrin repeat homology <PLK>
 F:221-269/Domain: SH3 homology <SH2>
 F:281-377/Domain: SH3 homology <SH2>
 F:400-658/Domain: protein kinase homology <KIN>
 F:408-416/Region: protein kinase ATP-binding motif

Query Match 12.9%; Score 117; DB 2; Length 659;
 Best Local Similarity 22.8%; Pred. No. 0.16;
 Matches 43; Conservative 32; Mismatches 66; Indels 46; Gaps 9;

OY 4 VYEKTP--QMGDPASLEPOIAETTSNIEKLE-VQKREA--WL-----AEAES---47
 DB 105 YDESEPLVVFSPTEELRKRMHQLKNVIRYNSDLYQKHPCFWIDGQYLCSCQAKNANG 164
 OY 48 -RVLSNRDLS-----RHARPPXPASAPPPSSSSASQDTKESSEEPSESDTPLY 101
 DB 165 COLLENKRGSLKPPGSSHKTKKPLPP-----TPEDQILKKPLPEAPPAAPV- 211
 OY 102 TEFDEFESEPTSPIGHCAVAYHFEESSEGTISMAEGEDLSMEEDKGDGWTVRVRKEGG 161
 DB 212 -----STSELKKVVALYDMPMANNDLQLRKGEYFLESNLP-WWRARDKNGQ 260
 OY 162 EGYVPTSL 170
 DB 261 EGYIPSNVY 269

RESULT 12

A34104
 Protein-tyrosine kinase (EC 2.7.1.112) src 1 [similarity] - African clawed frog
 N:Alternate names: kinase-related transforming protein (src); kinase-related transformi
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 16-Jun-2000 #sequence_revision 16-Jun-2000 #text_change 16-Jun-2000
 C:Accession: A34104; 151564
 R:Steele, R.E.; Unger, T.F.; Mardis, M.J.; Ferro, J.B.
 J. Biol. Chem. 264, 10649-10653, 1989
 A:Title: The two Xenopus laevis SRC genes are co-expressed and each produces functional
 A:Reference number: A34104; MUID:89278134
 A:Accession: A34104
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-532 <STE>
 A:Cross-references: GB:M24704; GB:J04822; NID:g214804; PIDN:AAA49962.1; PID:g214805
 R:Steele, R.E.; Chosin, R.; Ral, B.B.A.; Winkkur, S.T.; Unger, T.F.
 Oncogene 7, 2345-2350, 1992
 A:Title: Structural organization of a src gene from xenopus laevis.
 A:Reference number: 151564; MUID:93064714
 A:Accession: 151564
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-113 <STR2>
 A:Cross-references: GB:M33646; NID:g214808; PIDN:AAA49963.1; PID:g214810
 C:Genetics: 80/1
 A:Introns: 80/1
 C:Superfamily: protein-tyrosine kinase src; protein kinase homology; SH2 homology; SH3 h
 C:Keywords: ATP; autophosphorylation; blocked amino end; lipoprotein; myristylation; pho
 F:87-136/Domain: SH3 homology <SH2>
 F:147-244/Domain: SH2 homology <SH2>
 F:264-522/Domain: protein kinase homology <KIN>
 F:272-280/Region: protein kinase ATP-binding motif
 F:2/Modified site: myristylated amino end (Gly) (in mature form) #status predicted
 F:294/Active site: Lys #status predicted
 F:415,526/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status prec

Query Match 12.7%; Score 116; DB 1; Length 532;
 Best Local Similarity 29.6%; Pred. No. 0.14;
 Matches 32; Conservative 19; Mismatches 45; Indels 12; Gaps 3;
 OY 75 SSNSASQDTKES--SEEPSESDTPYTEDE-DFEETPSP-----IGHCAVAI 122
 DB 29 TSLASQTPKNSKLSHNPAPPGGNCULTPFGGINSFDTITSPORTGPLAGVTFVAL 88

OY 123 YHEGSEGTSMAGEDLSLMEEDKGDGWTVRVRKEGGYVPTSL 170
 DB 89 YVESRTETDLSFKGERLQIVNTEGDMWLARSLSSGGOTYIPSNVY 136

RESULT 13

B34104
 Protein-tyrosine kinase (EC 2.7.1.112) src 2 [similarity] - African clawed frog
 N:Alternate names: kinase-related transforming protein (src); kinase-related transfor
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 16-Jun-2000 #sequence_revision 16-Jun-2000 #text_change 16-Jun-2000
 C:Accession: B34104; 151563
 R:Steele, R.E.; Unger, T.F.; Mardis, M.J.; Ferro, J.B.
 J. Biol. Chem. 264, 10649-10653, 1989
 A:Title: The two Xenopus laevis SRC genes are co-expressed and each produces function
 A:Reference number: A34104; MUID:89278134
 A:Accession: B34104
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-532 <STE>
 A:Cross-references: GB:M23422; GB:J04822; NID:g214796; PIDN:AAA49961.1; PID:g214797
 R:Steele, R.E.
 Nucleic Acids Res. 13, 1747-1761, 1985
 A:Title: Two divergent cellular src genes are expressed in Xenopus laevis.
 A:Reference number: 151563; MUID:85215578
 A:Accession: 151563
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 439-492 <STR2>
 A:Cross-references: GB:M30858; NID:g214799; PIDN:AAA51644.1; PID:g555569
 C:Genetics:
 A:Gene: src
 C:Superfamily: protein-tyrosine kinase src; protein kinase homology; SH2 homology; SH
 C:Keywords: ATP; autophosphorylation; blocked amino end; lipoprotein; myristylation;
 F:87-136/Domain: SH3 homology <SH2>
 F:147-244/Domain: SH2 homology <SH2>
 F:264-522/Domain: protein kinase homology <KIN>
 F:272-280/Region: protein kinase ATP-binding motif
 F:2/Modified site: myristylated amino end (Gly) (in mature form) #status predicted
 F:294/Active site: Lys #status predicted
 F:415,526/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status p

Query Match 12.5%; Score 113.5; DB 1; Length 532;
 Best Local Similarity 26.0%; Pred. No. 0.22;
 Matches 34; Conservative 18; Mismatches 50; Indels 29; Gaps 4;

OY 69 SAPPPSSNSASQDTKESSEEP-PSRESDTPY-----YTF-DEDF 108
 DB 6 SKPREGGPRSKRLDIAEGSHQFTSLASQTPSKSLDSIRPGGPGGNCULTPPEGVAF 65
 OY 109 EEEPTSP-----IGHCAVAYHFEESSEGTISMAEGEDLSLMEEDKGDGWTVRKRE 159
 DB 66 SPTITSPQGTGLAGVTFVALYDESRTEITDLSFRKGERLQIVNTEGDMWLARSLSS 125
 OY 160 GGEYVPTSL 170
 DB 126 GQTYIPSNVY 136

RESULT 14

SS4570
 Probable membrane protein YMR109W - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein YMR109W
 C:Species: Saccharomyces cerevisiae
 C:Date: 08-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 02-Feb-2001
 C:Accession: SS4570
 R:Hunt, S.; Bowman, S.
 submitted to the EMBL Data Library, May 1995
 A:Reference number: SS4570
 A:Accession: SS4570

A:Molecule type: DNA
A:Residues: 1-1219 <HUN>
A:Cross-references: EMBL:249702; NID:9817859; PID:9817867; MIPS:YMR109w
A:Experimental source: strain AB972
C:Genetics:

A:Gene: SGD:MYO5
A:Cross-references: SGD:S0004715; MIPS:YMR109w
A:Map position: 13R

C:Superfamily: protozoan myosin heavy chain IB; myosin motor domain homology; SH3 homolo
C:Keywords: nucleotide binding; P-loop; transmembrane protein

F:39702/Domain: myosin motor domain homology <MMOT>

F:129-136/Region: nucleotide-binding motif A (P-loop)

F:293-309/Domain: transmembrane #status predicted <TM>

F:1092-1142/Domain: SH3 homology <SH3>

Query Match 12.4%; Score 112.5; DB 2; Length 1219;

Best Local Similarity 22.6%; Pred. No. 0.72;

Matches 48; Conservative 38; Mismatches 69; Indels 57; Gaps 11;

OY 3 DVEKTP---QMGDPASLE---POIAETLSNIERLKEVQYEAWLAEASRVLSNRGD 55

Db 946 DIVKSTISVVRGNPNPNSQVHKKPKKSSISS-----GYHASSQATRRPVSIATA 996

OY 56 S-----LSRHARPPXP-----ASAPDDSSNSASQDTKESSEPPSESDT 98

Db 997 QHVP7APASRHSKRPAPPPGMQNKATRRSVNPASTLTASQSNRPS--PPTAATRAT 1054

OY 99 PIYTFEDDF-----EEPTSPIGHCAIYHFEES-SEGTISMAEGED 140

Db 1055 PAATPAAAMSGRQANIPPPPPPPSSKPKPEMFE--AAVDEPGSGSPSELPLKKGDV 1112

OY 141 LSLMEDKGDGWTVRVRKEG-EGYVPTSYLR 171

Db 1113 IYITREP-SGMSLCKLDDGSKGWPPTAYMK 1143

RESULT 15

T39427

Probable myosin I protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 08-Sep-2000

C:Accession: T39427

R:McDougal, R.; Rajandream, M.A.; Barrel, B.G.; Ramsperger, U.; Bothe, G.; Pohl, T.

submitted to the EMBL data library, July 1999

A:Reference number: Z21853

A:Accession: T39427

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1217 <MCD>

A:Cross-references: EMBL:AL096797; PIDN:CAB46766.1; GSPDB:GN00067; SPDB:SPBC146.13c

A:Experimental source: strain 972h; cosmid c146

C:Genetics:

A:Gene: SPDB:SPBC146.13c

A:Map position: 2

A:Insertions: 1/3

C:Superfamily: protozoan myosin heavy chain IB; myosin motor domain homology; SH3 homolo
F:43-707/Domain: myosin motor domain homology <MMO>

Query Match 12.3%; Score 112; DB 2; Length 1217;

Best Local Similarity 23.7%; Pred. No. 0.78;

Matches 33; Conservative 27; Mismatches 57; Indels 22; Gaps 4;

OY 52 NRGDSLNRHARPPXPASAPDDSSNSASQDTKESSEPP-----PSESO-DTP 99

Db 1024 NRPPAPAMQARPNNTQAAPVSTTTIKQATTVTSASKAPASTVTSASSPSNISKPSAP 1083

OY 100 IYTFEDDFEETSPIGHC-----VAIYHFEESSEGTISMAEGEDLSLMEEDKGD 151

Db 1084 VANNVSKPSAVPPPPPAVEYKKDLVLAIDFAGRSNPENMTIKKDEILEIYQKERSGW 1143

OY 152 WTRVRRKKEGGGYVPTSYL 170

Db 1144 WLAL--KNGAEGWVPATYV 1160

Search completed: April 7, 2002, 16:09:15
Job time: 193 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2002, 16:14:11 ; Search time 15.46 Seconds

(without alignments)
415.029 Million cell updates/sec

Title: US-09-925-122A-3

Perfect score: 910
Sequence: 1 MKDYKTPQMGDPASLEPQ.....RRKGEGVPTSLRVTLN 175

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	538	59.1	115	TR10_HUMAN	Q15642 homo sapien
2	152	16.7	642	YB65_SCHPO	Q09746 schizosacch
3	130	14.3	633	YHR4_YEAST	P38822 saccharomyc
4	119	13.1	659	BTK_MOUSE	P35991 mus musculu
5	117	12.9	659	BTK_HUMAN	Q06187 mus sapien
6	116	12.7	531	SRCT_XENLA	P13115 xenopus lae
7	113.5	12.5	531	SRCT_XENLA	P13116 xenopus lae
8	112.5	12.4	1219	YMD2_YEAST	Q04439 saccharomyc
9	111	12.2	631	TEC_HUMAN	P42680 homo sapien
10	108.5	11.9	526	SRC_AVISR	P00525 avian sarco
11	108.5	11.9	1111	MYSH_DICDI	P34092 dictyosteli
12	108	11.9	532	SRC_CHICK	P00523 gallus gall
13	108	11.9	537	SRC_AVIST	P14085 avian sarco
14	108	11.9	568	SRC_AVIST	P14084 avian sarco
15	108	11.9	587	SRC_AVIS2	P15054 avian sarco
16	107.5	11.8	535	SRC_RAT	O9wud9 rattus norv
17	106.5	11.7	526	SRC_RSVPL	P00526 rous sarcom
18	106	11.6	526	SRC_RSVPL	P25020 rous sarcom
19	105	11.5	620	ITK_HUMAN	Q08881 homo sapien
20	104.5	11.4	535	TEC_MOUSE	P24604 mus musculu
21	104	11.4	535	SRC_HUMAN	P12931 homo sapien
22	104	11.4	541	YES_MOUSE	Q04736 mus musculu
23	103.5	11.4	1123	ABLI_MOUSE	P00520 mus musculu
24	103.5	11.4	1130	ABLI_HUMAN	P00519 homo sapien
25	103	11.3	536	FTN_XENLA	P13406 xenopus lae
26	102.5	11.3	450	CSK_CHICK	P41239 gallus gall
27	102.5	11.3	535	YRK_CHICK	Q02977 gallus gall
28	102.5	11.3	1182	ABLI_HUMAN	P42664 homo sapien
29	102.5	11.3	1232	B3A3_HUMAN	P48751 homo sapien
30	102	11.2	505	SRK1_SPOLA	P42686 spongilla 1
31	102	11.2	506	SRK1_SPOLA	P42680 spongilla 1
32	102	11.2	536	FTN_XIPHE	P27446 xiphophorus
33	101.5	11.2	450	CSK_HUMAN	P41240 homo sapien

34	101.5	11.2	450	1	CSK_MOUSE	P41241 mus musculu
35	101.5	11.2	450	1	CSK_RAT	P32577 rattus norv
36	101.5	11.2	671	1	VINE_HUMAN	Q06504 homo sapien
37	101	11.1	526	1	SRC_RSVSR	P00524 rous sarcom
38	100.5	11.0	527	1	TXK_MOUSE	P42682 mus musculu
39	100	11.0	821	1	EP8_MOUSE	Q08509 mus musculu
40	99.5	10.9	377	1	PEX8_HUMAN	Q75381 homo sapien
41	99.5	10.9	543	1	YES_HUMAN	P07947 homo sapien
42	99	10.9	376	1	PEXE_MOUSE	Q97080 mus musculu
43	98.5	10.8	546	1	SRC_MOUSE	Q06598 mus musculu
44	98.5	10.8	530	1	SRC_HUMAN	Q14247 homo sapien
45	98.5	10.8	1520	1	ABL_DROME	P00522 drosophila

ALIGNMENTS

RESULT	ID	TR10_HUMAN	STANDARD	PRT	115 AA
AC	Q15642				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	THYROID RECEPTOR INTERACTING PROTEIN 10 (TRIP10) (FRAGMENT).				
GN	TRIP10.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=95295737; PubMed=7776974;				
RA	Lee J.W., Choi H.-S., Gyuris J., Brent R., Moore D.D.;				
RT	"Two classes of proteins dependent on either the presence or absence				
RT	of thyroid hormone for interaction with the thyroid hormone				
RT	receptor.";				
RL	Mol. Endocrinol. 9:243-254(1995).				
CC	-1- FUNCTION: THYROID RECEPTOR INTERACTING PROTEINS (TRIPS)				
CC	SPECIFICALLY INTERACT WITH THE LIGAND BINDING DOMAIN OF THE				
CC	THYROID RECEPTOR (TR). TRIP10 REQUIRES THE PRESENCE OF THYROID				
CC	HORMONE FOR ITS INTERACTION.				
CC	-1- TISSUE SPECIFICITY: EXPRESSED MOST ABUNDANTLY IN SKELETAL MUSCLE,				
CC	HEART AND LUNG.				
CC	-1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
CC	EMBL; L40379; AAC41729.1; -				
CC	MTM; 604504; -				
DR	InterPro; IPR001452; SH3.				
DR	Pfam; PF00018; SH3; 1.				
DR	SMART; SM00326; SH3; 1.				
DR	PROSITE; PS50002; SH3; 1.				
KW	SH3 domain.				
FT	DOMAIN				
FT	60 115 SH3.				
FT	SEQUENCE 115 AA; 12848 MW; 54CD/DCDBFC10308 CRC64;				
QY	Query Match	59.1%;	Score 538;	DB 1;	Length 115;
QY	Best Local Similarity	99.0%;	Pred. No. 7.7e-33;		
QY	Matches 102;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;
DB	73 DSSNSASADTKESSEPEPSESODPIYTERDEDFEEPEPSPIGHCAVYHFESSSECT 132				
DB	13 NSSNSASODTKESSEPEPSESODPIYTERDEDFEEPEPSPIGHCAVYHFESSSECT 72				

OY 133 ISMAEGEDLSLMEDEKGDGWTNRVRKKEGGEGYPTSYLRVTLN 175
DB 73 ISMAEGEDLSLMEDEKGDGWTNRVRKKEGGEGYPTSYLRVTLN 115

RESULT 2

YB65_SCHPO STANDARD: PRT: 642 AA.

AC 009746:

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 72.2 KDA PROTEIN C12C2.05C IN CHROMOSOME II.
GN SPBC12C2.05C.

OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=972;
RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
BINDING DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
CC -1- SIMILARITY: TO YEAST YHR114W.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: Z54140; CAA90818.1; -.
DR HSSP: P29354; 1GFD.
DR InterPro: IPR002219; DAG_PE_bind.
DR InterPro: IPR001060; FCH.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00130; DAG_PE_bind; 1.
DR Pfam: PF00611; FCH; 1.
DR Pfam: PF00018; SH3; 2.
DR PRINTS: PR00008; DAGPEDOMAIN.
DR SMART: SM00109; C1; 1.
DR SMART: SM00055; FCH; 1.
DR SMART: SM00326; SH3; 2.
DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 1.
DR PROSITE: PS50081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE: PS50002; SH3; 2.
KW Hypothetical protein; SH3 domain; Zinc; Phorbol-ester binding.
FT DOMAIN 397 447 PHORBOL-ESTER AND DAG BINDING.
FT DOMAIN 521 575 SH3 1.
FT DOMAIN 584 642 SH3 2.
SQ SEQUENCE 642 AA; 72216 MW; E6770ED15BD73D3C CRC64;

Query Match 16.7%; Score 152; DB 1; Length 642;
Best Local Similarity 30.0%; Pred. No. 0.00045;
Matches 33; Conservative 27; Mismatches 42; Indels 8; Gaps 3;

OY 69 SAPDSSNSASODTKESSEPPSEESODPTIY-----TFEDDEFEETPTPIGHCAV 122
DB 465 TATDDFDASASSNAYEYRNTYTDMDSSSIYQTSLSNVKTEETTPAPASKVVG-VYL 523
OY 123 YHFGSSEGTISMAEGEDLSLMEDEKGDGWTNRVRKKEGGEGYPTSYLRV 172
DB 524 YDFTGEHGVITASEGQETFLLEPDDGSGWVRV-KIDGTGLIPASYVKL 572

RESULT 3

YHR4_YEAST STANDARD: PRT: 633 AA.

AC P38822;

DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 71.2 KDA PROTEIN IN ERK5-ORC6 INTERGENIC REGION.
GN YHR114W.

OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=5288C / AB972;
RX MEDLINE=94378003; PubMed-8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nham M., Rifkin L., Riles L., St Peter H., Trevisan E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT *Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII.";

RT Science 265:2077-2082(1994).
CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
CC -1- SIMILARITY: TO S.POMBE SPBC12C2.05C.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: U00059; AAB68850.1; -.
DR PIR: S48956; S48956.
DR SGD: S0001156; BZ21.
DR InterPro: IPR001060; FCH.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00611; FCH; 1.
DR Pfam: PF00018; SH3; 2.
DR SMART: SM00055; FCH; 1.
DR SMART: SM00326; SH3; 2.
DR PROSITE: PS50002; SH3; 2.
KW Hypothetical protein; SH3 domain.
FT DOMAIN 493 555 SH3 1.
FT DOMAIN 577 633 SH3 2.
SQ SEQUENCE 633 AA; 71171 MW; 5C73DAC69611B41 CRC64;

Query Match 14.3%; Score 130; DB 1; Length 633;
Best Local Similarity 26.4%; Pred. No. 0.017;
Matches 32; Conservative 27; Mismatches 34; Indels 28; Gaps 4;

OY 58 SRHARPPAPAPDSSNS-----ASODTKESSEPPSEESODPTIYTFEDDEFEETPT 113
DB 456 SRRTRGSAPNNAEGSDNNISIRTTSTNKKTYON-SSDDGKNKVLVYVYVKDD-- 511
OY 114 SPIGHCAVYHFGSSEGTISMAEGEDLSLMEDEKGDGWTNRVRKKEGGEGYPTSYLRV 172
DB 512 -----ITTPEDKISLVARDIGSGMTKINNNTGTGTGLVPTTYINI 552

OY 173 T 173
DB 553 S 553

RESULT 4
BTK_MOUSE

ID BTK_MOUSE STANDARD; PRT; 659 AA.
AC P35991; 061365;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TYROSINE-PROTEIN KINASE BTK (EC 2.7.1.112) (BRUTON'S TYROSINE KINASE)
DE (AGAMAGLOBLINEMIA TYROSINE KINASE) (ATK) (B CELL PROGENITOR
DE KINASE) (BTK) (KINASE EMB).
GN BTK OR BTK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-93145329; PubMed-8425221;
RA Tsukada S., Saffran D.C., Rawlings D.J., Parolini O., Allen R.C.,
RA Kilsak I., Sparks R.S., Kubagawa H., Mohandas T., Quan S.,
RA Belmont J.W., Cooper M.D., Conley M.E., Witte O.N.;
RT "Deficient expression of a B cell cytoplasmic tyrosine kinase in
human X-linked agammaglobulinemia."; Cell 72:279-290(1993).
RL [2]
RN SEQUENCE FROM N.A.
RX MEDLINE-93336578; PubMed-8476425;
RA Yamada N., Kawakami Y., Kimura H., Fukamachi H., Balser G.,
RA Altman A., Kato T., Inagaki Y., Kawakami T.;
RT "Structure and expression of novel protein-tyrosine kinases, Emb and
RT Emc, in hematopoietic cells."; Biochem. Biophys. Res. Commun. 192:231-240(1993).
RL [3]
RN SEQUENCE FROM N.A.
RX MEDLINE-95081608; PubMed-7989760;
RA Sideras P., Mueller S., Shields H., Jin H., Khan W.N., Nilsson L.,
RA Parkinson E., Thomas J.D., Branden L., Larsson I., Paul W.E.,
RA Rosen F.S., Alt F.W., Vetric D., Smith C.I.E., Xanthopoulos K.G.;
RT "Genomic organization of mouse and human Bruton's agammaglobulinemia
RT tyrosine kinase (Btk) loci."; J. Immunol. 153:5607-5617(1994).
RL [4]
RN SEQUENCE FROM N.A.
RX STRAIN-C129;
RA MEDLINE-95352959; PubMed-7626884;
RA Oelken J.C., Liu X., Lu J., Allen R.C., Muzny D., Belmont J.W.,
RA Gibbs R.A.;
RT "Sixty-nine kilobases of contiguous human genomic sequence containing
RT the alpha-galactosidase A and Bruton's tyrosine kinase loci."; Mamm. Genome 6:334-338(1995).
RL [5]
RN VARIANT XID CYS-28.
RX MEDLINE-93324903; PubMed-8332901;
RA Rawlings D.J., Saffran D.C., Tsukada S., Largespada D.A.,
RA Grimaldi J.C., Cohen L., Mohr R.N., Bazan J.F., Howard M.,
RA Copeland N.G., Jenkins N.A., Witte O.N.;
RT "Mutation of unique region of Bruton's tyrosine kinase in
RT immunodeficient XID mice."; Science 261:358-361(1993).
RL [6]
CC -1- FUNCTION: PLAYS A CRUCIAL ROLE IN B-CELL ONTOGENY.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -1- DISEASE: DEFECTS IN BTK ARE THE CAUSE OF MURINE X-LINKED
CC IMMUNODEFICIENCY (XID).
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC DOMAIN. BELONGS TO THE BTK SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L08967; AAA37316.1; -;
DR EMBL: L10627; -; NOT_ANNOTATED_CDS.
DR EMBL: L29788; AAA6994.1; -;
DR EMBL: U58105; AAB47246.1; -;
DR HSSP: Q06187; 1BTK.
DR MGD: MGI:88216; Btk.
DR InterPro: IPR001562; BTK.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR001452; SH3.
DR InterPro: IPR001245; Tyr_kin.
DR Pfam: PF00779; BTK.1.
DR Pfam: PF00169; PH.1.
DR Pfam: PF00069; pkinase.1.
DR Pfam: PF00017; SH2.1.
DR Pfam: PF00018; SH3.1.
DR PRINTS: PR00109; TYRKINASE.
DR PRINTS: PR00402; TECHRCDOMAIN.
DR PRINTS: PR00452; SH3DOMAIN.
DR SMART: SM00107; BTK.1.
DR SMART: SM00233; PH.1.
DR SMART: SM001849; PH.
DR SMART: SM00252; SH2.1.
DR SMART: SM00326; SH3.1.
DR SMART: SM00219; Tyrcg.1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP.1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR.1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.
DR PROSITE: PS50001; SH2.1.
DR PROSITE: PS50002; SH3.1.
DR PROSITE: PS50003; PH_DOMAIN.1.
DR TRANSFERASE; Tyrosine protein kinase; Phosphorylation;
DR ATP-binding; SH3 domain; SH2 domain; Disease mutation.
FT DOMAIN 3 133
FT DOMAIN 214 274
FT DOMAIN 281 377
FT DOMAIN 402 655
FT NP_BIND 408 416
FT BINDING 430 430
FT ACT_SITE 521 521
FT MOD_RES 551 551
FT VARIANT 28 28
FT CONFLICT 67 67
FT CONFLICT 123 123
FT CONFLICT 197 197
FT CONFLICT 450 450
SQ SEQUENCE 659 AA; 76437 MW; E502B798BC36E223 CRC64;
Query Match 13.1%; Score 119; DB 1; Length 659;
Best Local Similarity 23.3%; Pred. No. 0.11; Mismatches 68; Indels 46; Gaps 9;
Matches 44; Conservative 31;
QY 4 YEKTP--QMGPASLAEQIAETLSNTERKLE-VQKTEA--WL-----AEAES--- 47
DB 105 YDEGPLYVFSSTELRRKRWITQKLVIRYNSDIAQKHPCWIDGQYLCSQTKNMG 164
QY 48 -RVLSNRGDSLS-----RHAPPPAPASAPDSSNSASQDTKESSEPPSESQDTPLY 101
DB 165 COILENRNGSLKPGSSHKTKKPLPP-----TPEDQILKPLPPEPTAAPI- 211
QY 102 TFFDDFEEPEPTSPIGHCAVIAHFEQSSRGTISMAEGDLSLMEPDKGQGWTRVRKREG 161
DB 212 -----STTELKVVALLDYMPMANNDLQLRGEEYFTLESNLP-WWRARDKNGQ 260
QY 162 EGYVPTSYL 170
DB 261 EGYIPSNVI 269

RESULT 5
 BTK_HUMAN
 ID BTK_HUMAN STANDARD: PRT: 659 AA.
 AC 006187;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TYROSINE-PROTEIN KINASE BTK (EC 2.7.1.112) (BRUTON'S TYROSINE KINASE)
 DE (AGAMAGLOBLULINAEIA TYROSINE KINASE) (ATK) (B CELL PROGENITOR
 KINASE) (BRK).
 GN BTK OR ATK OR AGMY1 OR BPK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93140868; PubMed=8380905;
 RA Veltie D., Vorechovsky I., Sideras P., Holland J., Davies A.,
 RA Flinter F., Hammarstrom L., Kinon C., Levinsky R., Bobrow M.,
 RA Smith C.I.E., Bentley D.R.;
 RT "The gene involved in X-linked agammaglobulinemia is a member of the
 RT src family of protein-tyrosine kinases.";
 RL Nature 361:226-233(1993).
 RN [2]
 RP ERRATUM.
 RA Veltie D., Vorechovsky I., Sideras P., Holland J., Davies A.,
 RA Flinter F., Hammarstrom L., Kinon C., Levinsky R., Bobrow M.,
 RA Smith C.I.E., Bentley D.R.;
 RL Nature 364:362-362(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Blood;
 RX MEDLINE=94377492; PubMed=8090769;
 RA Ohta Y., Haire R.N., Litman R.T., Fu S.M., Nelson R.P., Kratz J.,
 RA Kornfeld S.J., la Morena M., Good R.A., Litman G.W., Kratz J.,
 RT "Genomic organization and structure of Bruton agammaglobulinemia
 RT tyrosine kinase: localization of mutations associated with varied
 RT clinical presentations and course in X chromosome-linked
 RT agammaglobulinemia.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9062-9066(1994).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95012452; PubMed=7927535;
 RA Rohrer J., Parolino O., Belmont J.W., Conley M.E.;
 RT "The genomic structure of human BTK, the defective gene in X-linked
 RT agammaglobulinemia.";
 RL Immunogenetics 40:319-324(1994).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95352959; PubMed=7626884;
 RA Oeltjen J.C., Liu X., Lu J., Allen R.C., Muzny D., Belmont J.W.,
 RA Gibbs R.A.;
 RT "Sixty-nine kilobases of contiguous human genomic sequence containing
 RT the alpha-galactosidase A and Bruton's tyrosine kinase loci.";
 RL Mamm. Genome 6:334-338(1995).
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Oeltjen J.C., Malley T.M., Muzny D.M., Miller W., Gibbs R.A.,
 RA Belmont J.W.;
 RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 1-442 FROM N.A.
 RX MEDLINE=93145329; PubMed=8425221;
 RA Tsukada S., Saifan D.C., Rawlings D.J., Parolini O., Allen R.C.,
 RA Kislak I., Sparkes R.S., Kubagawa H., Mohandas T., Quan S.,
 RA Belmont J.W., Cooper M.D., Conley M.E., Witte O.N.;
 RT "Deficient expression of a B cell cytoplasmic tyrosine kinase in
 RT human X-linked agammaglobulinemia.";
 RL Cell 72:279-290(1993).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 1-170.
 RX MEDLINE=97361821; PubMed=9218782;
 RA Hyoenen M., Saraste M.;
 RT "Structure of the PH domain and Btk motif from Bruton's tyrosine
 RT kinase: molecular explanations for X-linked agammaglobulinemia.";
 RL EMBO J. 16:3396-3404(1997).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-170.
 RX MEDLINE=99216540; PubMed=10196129;
 RA Baraldi E., Carugo K.D., Hyoenen M., Surdo P.L., Riley A.M.,
 RA Potter B.V.L., O'Brien R., Ladbury J.E., Saraste M.;
 RT "Structure of the PH domain from Bruton's tyrosine kinase in complex
 RT with inositol 1,3,4,5-tetrakisphosphate.";
 RL Structure 7:449-460(1999).
 RN [10]
 RP STRUCTURE BY NMR OF 209-275.
 RX MEDLINE=98153176; PubMed=9485443;
 RA Hansson H., Mattsson P.T., Allard P., Haapaniemi P., Vihinen M.,
 RA Smith C.I., Haerd T.;
 RT "Solution structure of the SH3 domain from Bruton's tyrosine kinase.";
 RL Biochemistry 37:2912-2924(1998).
 RN [11]
 RP REVIEW ON XLA VARIANTS.
 RX MEDLINE=96174621; PubMed=8594569;
 RA Vihinen M., Iwata T., Kinon C., Kwan S.-P., Ochs H.D.,
 RA Vorechovsky I., Smith C.I.E.;
 RT "Btkbase, mutation database for X-linked agammaglobulinemia (XLA).";
 RL Nucleic Acids Res. 24:160-165(1996).
 RN [12]
 RP REVIEW ON XLA VARIANTS.
 RX MEDLINE=97169387; PubMed=9016530;
 RA Vihinen M., Belohradsky B.H., Haire R.N., Holinski-Feder E.,
 RA Kwan S.-P., Lappalainen I., Leivaesalho H., Lester T., Meindl A.,
 RA Ochs H.D., Ollila J., Vorechovsky I., Weiss M., Smith C.I.E.;
 RT "Btkbase, mutation database for X-linked agammaglobulinemia (XLA).";
 RL Nucleic Acids Res. 25:166-171(1997).
 RN [13]
 RP VARIANTS XLA TRP-288; GLY-307; ASP-607 AND A 7 AMINO ACIDS INS.
 RX MEDLINE=94214481; PubMed=8162056;
 RA Bradley L.A.D., Sweatman A.K., Lovering R.C., Jones A.M., Morgan G.,
 RA Levinsky R.J., Kinon C.;
 RT "Mutation detection in the X-linked agammaglobulinemia gene, BTK,
 RT using single strand conformation polymorphism analysis.";
 RL Hum. Mol. Genet. 3:79-83(1994).
 RN [14]
 RP VARIANTS XLA HIS-28 AND TRP-288.
 RX MEDLINE=94214435; PubMed=8162018;
 RA de Weers M., Mensink R.G.J., Kraakman M.E.M., Schuurman R.K.B.,
 RA Hendriks R.W.;
 RT "Mutation analysis of the Bruton's tyrosine kinase gene in X-linked
 RT agammaglobulinemia: identification of a mutation which affects the
 RT same codon as is altered in immunodeficient xid mice.";
 RL Hum. Mol. Genet. 3:161-166(1994).
 RN [15]
 RP SEQUENCE FROM N.A., AND VARIANTS XLA S-334; R-506; Q-520; W-562; K-630.
 RX MEDLINE=95152493; PubMed=7880320;
 RA Hagemann T.L., Chen Y., Rosen F.S., Kwan S.-P.;
 RT "Genomic organization of the Btk gene and exon scanning for mutations
 RT in patients with X-linked agammaglobulinemia.";
 RL Hum. Mol. Genet. 3:1743-1749(1994).
 RN [16]
 RP VARIANTS XLA D-113; C-361; Q-520; P-542; W-562; K-630 AND P-652.
 RX MEDLINE=95152494; PubMed=7849697;
 RA Conley M.E., Fitch-Hilgenberg M.E., Cleveland J.L., Parolini O.,
 RA Rohrer J.;
 RT "Screening of genomic DNA to identify mutations in the gene for
 RT Bruton's tyrosine kinase.";
 RL Hum. Mol. Genet. 3:1751-1756(1994).
 RN [17]
 RP VARIANTS XLA H-28; P-33; P-408; G-589; D-613 AND Q-260--E-280 DEL.
 RX MEDLINE=95152522; PubMed=7849721;
 RA Zhu O., Zhang M., Winkelstein J., Chen S.-H., Ochs H.D.;
 RT "Unique mutations of Bruton's tyrosine kinase in fourteen unrelated
 RT X-linked agammaglobulinemia families.";

RL Hum. Mol. Genet. 3:1899-1900(1994).
 RN [18]
 RP VARIANTS XLA E-430; O-520; O-525; P-562; V-582; G-589; E-594 & D-613.
 RX MEDLINE-95108046; PubMed-7809124;
 RA Vilhinen M., Vetrle D., Manier H.S., Ochs H.D., Zhu Q., Vorechovsky I.,
 Webster A.D.B., Notarangelo L.D., Nilsson L., Sowadski J.M.,
 Smith C.I.E.;
 RT "Structural basis for chromosome X-linked agammaglobulinemia: a
 tyrosine kinase disease";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:12803-12807(1994).
 RN [19]
 RP VARIANT XLA PHE-64, AND CHARACTERIZATION OF OTHER XLA VARIANTS.
 RX MEDLINE-95151728; PubMed-7849006;
 RA Vilhinen M., Zvelebil J.J.M., Zhu Q., Brodmann R.A., Ochs H.D.,
 Zegers B.J.M., Nilsson L., Waterfield M.D., Smith C.I.E.;
 RT "Structural basis for pleckstrin homology domain mutations in
 X-linked agammaglobulinemia";
 RL Biochemistry 34:1475-1481(1995).
 RN [20]
 RP VARIANTS XLA S-25; W-288; M-370; V-509; P-525; K-526; W-562; V-582; R-594.
 RX MEDLINE-95227177; PubMed-7711734;
 RA Vorechovsky I., Vilhinen M., de Saint Basile G., Honsova S.,
 Hammarstrom L., Mueller S., Nilsson L., Fischer A., Smith C.I.E.;
 RT DNA-based mutation analysis of Bruton's tyrosine kinase gene in
 patients with X-linked agammaglobulinemia";
 RL Hum. Mol. Genet. 4:51-58(1995).
 RN [21]
 RP VARIANTS XLA LYS-567; LEU-587 AND HIS-641.
 RX MEDLINE-95359977; PubMed-7633420;
 RA Jin H., Webster A.D.B., Vilhinen M., Sideras P., Vorechovsky I.,
 Hammarstrom L., Bernatowska-Matuszkiewicz E., Smith C.I.E.,
 Bobrow M., Vetrle D.;
 RT "Identification of Btk mutations in 20 unrelated patients with
 X-linked agammaglobulinemia (XLA)";
 RL Hum. Mol. Genet. 4:693-700(1995).
 RN [22]
 RP VARIANTS XLA PRO-33; GLN-520; CYS-641 AND GLY-302 DEL.
 RX MEDLINE-95359987; PubMed-7633429;
 RA Gaspar H.B., Bradley L.A.D., Katz F., Lovering R.C., Roifman C.M.,
 Morgan G., Levinsky R.J., Kinon C.;
 RT Mutation analysis in Bruton's tyrosine kinase, the X-linked
 agammaglobulinemia gene, including identification of an insertional
 hotspot";
 RL Hum. Mol. Genet. 4:755-757(1995).
 RN [23]
 RP VARIANTS XLA ASN-429 AND ARG-477.
 RX MEDLINE-96177680; PubMed-8634718;
 RA Vorechovsky I., Luo L., de Saint Basile G., Hammarstrom L.,
 Webster A.D.B., Smith C.I.E.;
 RT "Improved oligonucleotide primer set for molecular diagnosis of
 X-linked agammaglobulinemia: predominance of amino acid
 substitutions in the catalytic domain of Bruton's tyrosine kinase";
 RL Hum. Mol. Genet. 4:2403-2405(1995).
 RN [24]
 RP VARIANTS XLA GLU-302 AND ASP-476.
 RX MEDLINE-95353280; PubMed-7627183;
 RA Hagemann T.L., Rosen F.S., Kwan S.-P.;
 Query Match 12.9%; Score 117; DB 1; Length 659;
 Best Local Similarity 22.8%; Pred. No. 0.16; Indels 46; Gaps 9;
 Matches 43; Conservative 32; Mismatches 68;
 Oy 4 VYETTP--OMGDPALEFQIAETLSIERKLE-VOKYEA--WL-----AEAES---47
 Db 105 VYDGPPLVFEPTBELKRWIHOAKNVRYSNDLVQKYPFGWIGQYVLCSSQTKNMG 164
 Oy 48 -RVLSNRGDS-----RHAPRPAPASAPDSSNSASQTKESSEPPSESDTPIY 101
 Db 165 CQILENRGSLKPPSSHRTKPLP-----TPEDQILKRPDPPEPAAYV- 211
 Oy 102 TEPEDFEPTSPIGHCVATVHFEGSEGTISMAEGEDLSIMEDKGDGWTVRKKEG 161
 Db 212 -----STSELKVVALLDYMPNANDLOLAKKGDEYFLIESNP-WRRADKNGQ 260

Oy 162 EGYPTSYL 170
 Db 261 EGYPTSYL 269
 RESULT 6
 ID SRC1_XENLA STANDARD; PRT; 531 AA.
 AC P3115;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE TYROSIENE-PROTEIN KINASE SRC-1 (BC 2.7.1.112) (P60-SRC-1).
 GN SRC-1.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-89278134; PubMed-2499582;
 RA Steele R.E., Unger T.F., Mardis M.J., Fero J.B.;
 RT "The two Xenopus laevis SRC genes are co-expressed and each produces
 functional pp60src";
 RL J. Biol. Chem. 264:10649-10653(1989).
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
 PROTEIN TYROSINE PHOSPHATE.
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SH DOMAIN.
 CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
 DOMAIN, BELONGS TO THE SRC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC The European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M24704; AAA49962.1; -
 DR PIR; A34104; A34104.
 DR HSSP; P00523; 2PTK.
 DR InterPro: IPR000719; Euk_Pkinase.
 DR InterPro: IPR000980; SH2.
 DR InterPro: IPR001452; SH3.
 DR InterPro: IPR001245; Tyr_Kin.
 DR Pfam; PF00069; Pkinase; 1.
 DR Pfam; PF00017; SH2; 1.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00109; TYRKINASE.
 DR PRINTS; PR00401; SH2DOMAIN.
 DR PRINTS; PR00452; SH3DOMAIN.
 DR SMART; SM00252; SH2; 1.
 DR SMART; SM00326; SH3; 1.
 DR SMART; SM00219; Tyrc; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS50001; SH2; 1.
 DR PROSITE; PS50002; SH3; 1.
 DR TRANSFERASE; ATP-binding; Tyrosine-protein kinase; Phosphorylation;
 KW Myristate; SH3 domain; SH2 domain.
 FT INIT MET 0
 FT LIPID 1 1
 FT DOMAIN 79 140 SH3.
 FT DOMAIN 146 243 SH2.
 FT DOMAIN 265 518 PROTEIN KINASE.
 FT NP_BIND 271 279 ATP (BY SIMILARITY).
 FT BINDING 293 293 ATP (BY SIMILARITY).
 FT ACT_SITE 384 384 BY SIMILARITY.


```

DR InterPro: IPR001452; SH3.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head.1.
DR Pfam: PF00018; SH3.1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRINTS: PR00452; SH3DOMAIN.
DR PRODOM: PD000355; myosin_head.1.
DR SMART: SM00242; MYSC.1.
DR SMART: SM00326; SH3.1.
DR PROSITE: PS50002; SH3.1.
DR Hypothetical protein; Myosin; Coiled coil; ATP-binding; SH3 domain.
KM DOMAIN 1 MYOSIN HEAD-LIKE.
FT DOMAIN 1085 1147 SH3.
FT DOMAIN 1190 1218 COILED COIL (POTENTIAL).
FT NP_BIND 129 136 ATP (POTENTIAL).
FT DOMAIN 1011 1016 POLY-PRO.
FT DOMAIN 1060 1063 POLY-ALA.
FT DOMAIN 1073 1081 POLY-PRO.
FT DOMAIN 1204 1218 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 1219 AA; 136898 MW; DFE9EC16B61CD29 CRC64;

Query Match 12.4%; Score 112.5; DB 1; Length 1219;
Best Local Similarity 22.6%; Pred. No. 0.67;
Matches 48; Conservative 38; Mismatches 69; Indels 57; Gaps 11;

OY 3 DYKETP---QMGDPASLE---POLAETLSNIERLKVQKYEAMLAESRVLISNRGD 55
DB 946 DYKSTTSVRKGNPNNSOVHKKPRKSSISS-----GYHSSSQATRPPIAANA 996
OY 56 S-----LSRHARPPXP-----ASAPDSSNSASQDTRESSEPPSESDPT 98
DB 997 QHVTPAPASRSHKAPPPPPQMNKATRRSVNPASTITASQSNARPS--PPTATRAIT 1054
OY 99 PIYTFDEDF-----EEPTSPIGHCAVTHFEGS--SEGITSMAGGD 140
DB 1055 PAATPAAAMGSGRGQANIPPPPPPPSSKPKPEPME--AAVDPSGSPSELPLKKGDV 1112
OY 141 LSLMEEDKGGDWTRVRRKGG--EGYVPTSYLR 171
DB 1113 IYITREEP--SGHSLGKLLDGSKEGWVPARYMK 1143

RESULT 9
TEC_HUMAN STANDARD; PRT; 631 AA.
AC P42680;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE TYROSINE-PROTEIN KINASE TEC (EC 2.7.1.112).
GN TEC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RA MEDLINE=95019607; PubMed=7934162;
RA Sato K., Mano H., Ariyama T., Inazawa J., Yazaki Y., Hirai H.;
RT "Molecular cloning and analysis of the human Tec protein-tyrosine
RT kinase.";
RL Leukemia 8:1663-1672(1994).
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -1- TISSUE SPECIFICITY: HEMATOPOIETIC CELL LINES INCLUDING MYELOID,
CC B-, AND T-CELL LINEAGES.
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC

```

```

CC DOMAIN: BELONGS TO THE BTK SUBFAMILY.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D29767; BAA06171.1; -.
DR HSSP: Q06187; 1BTK.
DR MIM: 600583; -.
DR InterPro: IPR001562; BTK.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000980; SH2.
DR InterPro: IPR001452; SH3.
DR InterPro: IPR001245; Tyr_kin.
DR Pfam: PF00779; BTK.1.
DR Pfam: PF00169; PH.1.
DR Pfam: PF00069; pkinase.1.
DR Pfam: PF00017; SH2.1.
DR Pfam: PF00018; SH3.1.
DR PRINTS: PR00109; TYRKINASE.
DR PRINTS: PR00402; TECBTKDOMAIN.
DR PRINTS: PR00452; SH3DOMAIN.
DR SMART: SM00107; BTK.1.
DR SMART: SM00233; PH.1.
DR SMART: SM00252; SH2.1.
DR SMART: SM00326; SH3.1.
DR SMART: SM00219; TYRK.1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP.1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR.1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM.1.
DR PROSITE: PS50001; SH2.1.
DR PROSITE: PS50002; SH3.1.
DR PROSITE: PS50003; PH_DOMAIN.1.
DR Transferrase; Tyrosine-protein kinase; ATP-binding; SH2 domain;
KM SH3 domain; Phosphorylation.
KW SH3 DOMAIN 4 111 PH.
FT DOMAIN 179 239 SH3.
FT DOMAIN 247 345 SH2.
FT DOMAIN 370 623 PROTEIN KINASE.
FT NP_BIND 376 384 ATP (BY SIMILARITY).
FT BINDING 398 398 ATP (BY SIMILARITY).
FT ACT_SITE 489 489 BY SIMILARITY.
FT MOD_RES 519 519 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 631 AA; 73629 MW; A5DDECAF991A9022 CRC64;

Query Match 12.2%; Score 111; DB 1; Length 631;
Best Local Similarity 26.6%; Pred. No. 0.41;
Matches 38; Conservative 20; Mismatches 39; Indels 46; Gaps 7;

OY 29 ERKLEVOXYEAMLAESAERVLISNRGDSLRHARPPAPASAPDSSNSASQDTRESSE 88
DB 137 EKLAPGCEK-----NLFESSTRKALPPAP-----ETKKRP 168
OY 89 EPPSESDPTIYTERDEDEFEETPSPIGHCAVTHFEGSSEGTISMAGEEDLSLMEEDK 148
DB 169 PEP-----IPLE---EDNSEIIV-----VAMYDQAAEGHDLRLRNGQEVYLE--K 211
OY 149 GD-GWTRVRRKKEGEGYVPTSYL 170
DB 212 NDVHWRRARDKYNGEYIPSNV 234

RESULT 10
SRC_AVISR
ID SRC_AVISR STANDARD; PRT; 526 AA.
AC P00525;

```

DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE TYROSINE-PROTEIN KINASE TRANSFORMING PROTEIN SRC (EC 2.7.1.112) (P60-SRC).
 GN V-SRC.
 OS Avian sarcoma virus (strain rASV1441).
 OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
 OX NCBI_TaxID=11894;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83059858; PubMed=6292477;
 RA Takeya T., Feldman R.A., Hanatusa H.;
 RT "DNA sequence of the viral and cellular src gene of chickens. 1.
 RT Complete nucleotide sequence of an EcoRI fragment of recovered avian
 RT sarcoma virus which codes for gp37 and pp60src.";
 RL J. Virol. 44:11(1982).
 RN [2]
 RP PHOSPHORYLATION AT TYR-416.
 RX MEDLINE=81220979; PubMed=6264320;
 RA Neil J.C., Ghysdael J., Vogt P.K., Smart J.E.;
 RT "Homologous tyrosine phosphorylation sites in transformation-specific
 RT gene products of distinct avian sarcoma viruses.";
 RL Nature 291:675-677(1981).
 CC -1- FUNCTION: THIS PHOSPHOPROTEIN, REQUIRED FOR BOTH THE INITIATION
 CC AND THE MAINTENANCE OF NEOPLASTIC TRANSFORMATION, IS A PROTEIN
 CC KINASE THAT CATALYZES THE PHOSPHORYLATION OF TYROSINE RESIDUES
 CC IN VITRO.
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
 CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC -----
 DR EMBL, K00928; AAA42565.1; -.
 DR PIR, A00631; TVEV60.
 DR HSSP, P00523; 2PTK.
 DR InterPro: IPR000719; Euk_kinase.
 DR InterPro: IPR000980; SH2.
 DR InterPro: IPR001452; SH3.
 DR InterPro: IPR001245; Tyr_kin.
 DR Pfam, PF00069; Pkinase; 1.
 DR Pfam, PF00017; SH2; 1.
 DR Pfam, PF00018; SH3; 1.
 DR PRINTS: PRO0109; TYRKINASE.
 DR PRINTS: PRO0401; SH2DOMAIN.
 DR PRINTS: PRO0452; SH3DOMAIN.
 DR SMART, SM00252; SH2; 1.
 DR SMART, SM00326; SH3; 1.
 DR SMART, SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE: PS50001; SH2; 1.
 DR PROSITE: PS50002; SH3; 1.
 DR Tyrosine-protein kinase; Transforming protein; Oncogene;
 KW Transferase; Phosphorylation; ATP-binding; Myristate;
 KM SH3 domain; SH2 domain.
 FT LIPID 2
 FT 81 142 MYRISTATE.
 FT DOMAIN 148 245 SH2.
 FT DOMAIN 267 517 PROTEIN KINASE.
 FT NP_BIND 273 281 ATP (BY SIMILARITY).
 FT BINDING 295 295 ATP (BY SIMILARITY).

FT ACT SITE 386 386 BY SIMILARITY.
 FT MOD_RES 416 416 PHOSPHORYLATION (AUTO-).
 SQ SEQUENCE 526 AA; 58878 MW; 7DB3903F80233E49 CRC64;
 Query Match 11.9%; Score 108.5; DB 1; Length 526;
 Best Local Similarity 23.9%; Pred. No. 0.5;
 Matches 33; Conservative 25; Mismatches 57; Indels 23; Gaps 4;
 QY 54 GDSLSRHARPPAPAS-APPDSS-----SNSASQDTKESSEPPSESDPTLY 101
 DB 2 GSSKSKPDPSPQRCGLPEPDSYHNGFPASQTPNKTAAADTHRTPSRSGVATPEKLF 61
 QY 102 TEPDECFEPEPSP-----IGHCVAIYHEPGESEGTSAAGEEDSLMEDEKDDM 152
 DB 62 GGFNT--SDPTVSPQNALAGAVTTFVALDYESTRTDLSEKKGRIQIVNTEGDWM 119
 QY 153 TVRRKREGGEVVPISYL 170
 DB 120 LMSLTGTGTIPSNVY 137
 RESULT 11
 MSB_DICDI STANDARD; PRT; 1111 AA.
 AC P34092;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MYOSIN IB HEAVY CHAIN.
 GN MYOA OR DMIB.
 OS Dictyostellum discoideum (slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostellum.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX3.
 RX MEDLINE=89345628; PubMed=2762320;
 RA Jung G., Saxe C.L. III, Kimmel A.R., Hammer J.A. III;
 RT "Dictyostellum discoideum contains a gene encoding a myosin I heavy
 RT chain.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:6186-6190(1989).
 RN [2]
 RP SEQUENCE OF 481-490; 656-666 AND 783-798.
 RC STRAIN=AX3;
 RX MEDLINE=93315475; PubMed=8325874;
 RA Jung G., Fukui Y., Martin B., Hammer J.A. III;
 RT "Sequence, expression pattern, intracellular localization, and
 RT targeted disruption of the Dictyostellum myosin I heavy chain
 RT isoform.";
 RL J. Biol. Chem. 268:14981-14990(1993).
 CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
 CC ACTIVITY THAT IS ACTIVATED BY ACTIN. MYOSIN IB MAY HAVE A ROLE IN
 CC CHEMOTAXIS AND AGGREGATION. IT COULD SERVE TO STABILIZE AND EVEN
 CC RETRACT CORTICAL STRUCTURES, SUCH AS PSEUDOPODS AND LAMELLOPODS.
 CC -1- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
 CC AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
 CC -1- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION JUST BENEATH THE
 CC PLASMA MEMBRANE IN THE ANTERIOR PSEUDOPOD AT THE LEADING EDGE OF
 CC THE CELL.
 CC -1- DOMAIN: TH.1 BINDS DIRECTLY TO ANIONIC PHOSPHOLIPID MEMBRANES;
 CC MYOSINS I COULD THEREFORE MOVE ACTIN RELATIVE TO MEMBRANES AND
 CC VICE VERSA. TH.2 AND SH3 BIND TIGHTLY TO F-ACTIN; THIS TOGETHER
 CC WITH THE NUCLEOTIDE-SENSITIVE SITE IN THE HEAD, ALLOWS SINGLE
 CC MOLECULES OF MYOSIN I TO CROSS-LINK ACTIN FILAMENTS.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

entitles requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL, M26037; AAA33229.1; .

CC PIR: A33284; A33284.

DR HSSP: P08799; 1MND.

DR Dictydb; DD01047; myoa.

DR InterPro; IPR001452; SH3.

DR InterPro; IPR001609; myosin_head.

DR Pfam; PF00063; myosin_head; 1.

DR Pfam; PF00018; SH3; 1.

DR PRINTS; PR00193; MYOSINHEAVY.

DR PRINTS; PR00452; SH3DOMAIN.

DR PRODOM; PD000355; myosin_head; 1.

DR SMART; SM00242; MYSC; 1.

DR SMART; SM00326; SH3; 1.

DR PROSITE; PS00022; SH3; 1.

DR Myosin; Actin-binding; ATP-binding; SH3 domain; Multigene family;

KW Chemotaxis; Phosphorylation.

KW MYOSIN HEAD-LIKE.

FT DOMAIN 1 694

FT DOMAIN 695 921

FT DOMAIN 922 1052

FT DOMAIN 1053 1111

FT NP BIND 102 109

FT DOMAIN 547 627

FT DOMAIN 951 1015

FT MOD_RES 332 332

FT SEQUENCE 1111 AA: 124313 MW: 65386 F08DC5642F CRC64;

Query Match 11.9%; Score 108.5; DB 1; Length 1111;

Best Local Similarity 34.4%; Pred. No. 1.2; Indels 3; Gaps 2;

Matches 21; Conservative 17; Mismatches 20; Indels 3; Gaps 2;

QY 111 EPTSPICHCVAIYFEGSSSEGTISMAEGEDLSMEDKDGWTRVRKKGEGYPTSYL 170

DB 1052 QPRLPRAK--ALYDYDASSSTDELSPFKEDGDIIFVCKDNG-GWTGGLKSGQKQWPTNTYL 1108

QY 171 R 171

DB 1109 Q 1109

RESULT 12

SRC_CHICK STANDARD; PRT; 532 AA.

AC P00523; Q91345; Q92013; Q98915; Q91343; Q90992;

DT 21-JUL-1986 (Rel. 01, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE SRC (EC 2.7.1.112) (P60-SRC)

DE (C-SRC).

GN SRC.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OC NCBI_TaxID=9031;

OX [1]

RX MEDLINE=8315564; PubMed=6299580;

RX MEDLINE=8315564; PubMed=6299580;

RA Takeya T., Hanafusa H.;

RT "Structure and sequence of the cellular gene homologous to the RSV src gene and the mechanism for generating the transforming virus.";

RL Cell 32:881-890(1983).

RL [2]

RN REVISION TO 525.

RA Takeya T., Hanafusa H.;

RL Cell 34:319-319(1983).

RL [3]

RN SEQUENCE FROM N.A. AND PHOSPHORYLATION AT TYR-415 AND TYR-435.

RX MEDLINE=97008971; PubMed=8856081;

RA Weijland A., Neubauer G., Courtneidge S.A., Mann M., Wierenga R.K.,

RA Superli-Furga G.;

RT "The purification and characterization of the catalytic domain of Src

RT expressed in Schizosaccharomyces pombe. Comparison of

RT unphosphorylated and tyrosine phosphorylated species.";

RL Eur. J. Biochem. 240:756-764(1996).

RL [4]

RN SEQUENCE OF 1-18 AND 484-533 FROM N.A.

RX MEDLINE=91304409; PubMed=1712905;

RA Dorai T., Levy J.B., Kang L., Brugge J.S., Wang L.H.;

RT "Analysis of cDNAs of the proto-oncogene c-src: heterogeneity in 5'

RT exons and possible mechanism for the genesis of the 3' end of

RT v-src.";

RL Mol. Cell. Biol. 11:4165-4176(1991).

RL [5]

RN ATP-BINDING SITE.

RX MEDLINE=84270751; PubMed=6431300;

RA Kamps M.P., Taylor S.S., Sefton B.M.;

RT "Direct evidence that oncogenic tyrosine kinases and cyclic AMP-

RT dependent protein kinase have homologous ATP-binding sites.";

RL Nature 310:589-592(1984).

RL [6]

RN PHOSPHORYLATION.

RX MEDLINE=86028181; PubMed=2996780;

RA Gould K.L., Woodgett J.R., Cooper J.A., Buss J.E., Shalloway D.,

RA Hunter T.;

RT "Protein kinase C phosphorylates pp60src at a novel site.";

RL Cell 42:849-857(1985).

RL [7]

RN PHOSPHORYLATION AT TYR-415.

RX MEDLINE=82082387; PubMed=6273838;

RA Smart J.E., Oppermann H., Czernilofsky A.P., Purchio A.F.,

RA Erikson R.L., Bishop J.M.;

RT "Characterization of sites for tyrosine phosphorylation in the

RT transforming protein of Rous sarcoma virus (pp60v-src) and its normal

RT cellular homologue (pp60c-src).";

RL Proc. Natl. Acad. Sci. U.S.A. 78:6013-6017(1981).

RL [8]

RN PHOSPHORYLATION AT TYR-526.

RX MEDLINE=86515652; PubMed=2420005;

RA Cooper J.A., Gould K.L., Cartwright C.A., Hunter T.;

RT "Ty527 is phosphorylated in pp60c-src: implications for regulation.";

RL Science 231:1431-1434(1986).

RL [9]

RN X-RAY CRYSTALLOGRAPHY (2.35 ANGSTROMS) OF 82-532.

RX MEDLINE=98070614; PubMed=9405157;

RA Williams J.C., Weijland A., Gonfalon S., Thompson A.,

RA Courtneidge S.A., Superli-Furga G., Wierenga R.K.;

RT "The 2.35 A crystal structure of the inactivated form of chicken Src:

RT a dynamic molecule with multiple regulatory interactions.";

RL J. Mol. Biol. 274:757-775(1997).

RL [10]

RN STRUCTURE BY NMR OF 80-139.

RX MEDLINE=93279385; PubMed=8504863;

RA Yu H., Rosen M.K., Schneider S.L.;

RT "1H and 15N assignments and secondary structure of the Src SH3

RT domain.";

RL FEBS Lett. 324:87-92(1993).

RL [11]

RN STRUCTURE BY NMR OF 76-139.

RX MEDLINE=95063992; PubMed=7526465;

RA Feng S., Chen J.K., Yu H., Simon J.A., Schneider S.L.;

RT "Two binding orientations for peptides to the Src SH3 domain:

RT development of a general model for SH3-ligand interactions.";

RL Science 266:1241-1247(1994).

RL [12]

RN FUNCTION: THE FUNCTION OF PP60-C-SRC IS UNKNOWN. IT IS EXPRESSED

CC TO HIGH LEVELS, AND WITH A HIGH DEGREE OF KINASE ACTIVITY, IN

CC CERTAIN FULLY DIFFERENTIATED CELLS SUCH AS NEURONS, PLATELETS

CC AND MACROPHAGES.

CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +

CC PROTEIN TYROSINE PHOSPHATE.

CC -1- ENZYME REGULATION: BECOMES ACTIVATED WHEN ITS MAJOR TYROSINE

CC PHOSPHORYLATION SITE IS NOT PHOSPHORYLATED. IT CAN ALSO BE

CC ACTIVATED BY POINT MUTATIONS AS WELL AS BY TRUNCATIONS AT THE

CC C-TERMINAL END OR BY OTHER MUTATIONS.
 CC -1- PTM: PHOSPHORYLATED BY C-SRC KINASE (CSK) IN THE C-TERMINUS. THE
 CC PHOSPHORYLATED TAIL INTERACTS WITH THE SH2 DOMAIN THEREBY
 CC REPRESSING KINASE ACTIVITY (BY SIMILARITY).
 CC -1- MISCELLANEOUS: POLYOMA VIRUS MIDDLE T ANTIGEN FORMS A COMPLEX WITH
 CC P60-C-SRC.
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
 CC DOMAIN, BELONGS TO THE SRC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: V00402; CAA23696.1; -
 CC EMBL: J00844; AAA70194.1; -
 CC EMBL: S43604; AAD13831.1; -
 CC EMBL: S43616; AAD13835.1; -
 CC EMBL: S43587; AAD13830.1; -
 CC EMBL: S43609; AAD13832.1; -
 CC EMBL: S43614; AAD13834.1; -
 CC EMBL: S43579; AAB19353.2; -
 CC PIR: A00630; TVCHS.
 CC PDB: 1SRJ; 31-MAY-94.
 CC PDB: 1SRM; 31-MAY-94.
 CC PDB: 2PTK; 24-DEC-97.
 CC PDB: 1QWE; 08-MAR-96.
 CC PDB: 1QWE; 08-MAR-96.
 CC PDB: 1PRL; 07-FEB-95.
 CC PDB: 1PRM; 07-FEB-95.
 CC PDB: 1RLP; 07-FEB-95.
 CC PDB: 1RLQ; 07-FEB-95.
 CC PDB: 1RLO; 27-JAN-97.
 CC PDB: 1NLP; 27-JAN-97.
 CC InterPro: IPR000719; Euk_Pkinase.
 CC InterPro: IPR000980; SH2.
 CC InterPro: IPR001452; SH3.
 CC InterPro: IPR001245; Tyr_kin.
 CC Pfam: PF00069; pkinase; 1.
 CC Pfam: PF00017; SH2; 1.
 CC Pfam: PF00018; SH3; 1.
 CC PRINTS: PR00109; TYRKINASE.
 CC PRINTS: PR00401; SH2DOMAIN.
 CC PRINTS: PR00452; SH3DOMAIN.
 CC SMART: SM00252; SH2; 1.
 CC SMART: SM00326; SH3; 1.
 CC SMART: SM00219; TYRK; 1.
 CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 CC PROSITE: PS50001; SH2; 1.
 CC PROSITE: PS50002; SH3; 1.
 CC Tyrosine-protein kinase; Proto-oncogene; Phosphorylation;
 CC Transferrase; ATP-binding; Myristate; SH3 domain; SH2 domain;
 CC 3D-structure.
 CC INIT_MET 0
 CC LIPID 1
 CC DOMAIN 80 141 MYRISTATE.
 CC DOMAIN 147 244 SH3.
 CC DOMAIN 266 519 SH2.
 CC MOD_RES 11 11 PROTEIN KINASE.
 CC NP_BIND 272 280 PHOSPHORYLATION (BY PKC).
 CC BINDING 294 294 ATP (BY SIMILARITY).
 CC ACT_SITE 385 385 BY SIMILARITY.
 CC MOD_RES 415 415 PHOSPHORYLATION (AUTO-).
 CC MOD_RES 435 435 PHOSPHORYLATION (AUTO-).
 CC MOD_RES 526 526 PHOSPHORYLATION (BY CSK; NEGATIVE
 CC REGULATION).

FT CONFLICT 300 300 T -> N (IN REF. 1).
 FT CONFLICT 500 500 K -> R (IN REF. 1).
 SQ SEQUENCE 532 AA; 59878 MW; 86DB036F6994E401 CRC64;
 Query Match 11.9%; Score 108; DB 1; Length 532;
 Best local Similarity 22.6%; Pred. No. 0.55; Mismatches 26; Gaps 4;
 Matches 31; Conservative 27; Mismatches 53; Indels 26; Gaps 4;
 QY 58 SRIARPPXP-----ASAPPDS-----SNSAQDTKESSEPPSESDPIYR 102
 Db 2 SSKSKPKDSQRRSLPPDTHHGGEPPASQTPNKTAAPDHTPERSFGVATEKRLG 61
 QY 103 EPFDEDEEPTSP-----IGHCAIYHFGSGSEGTISMAEGEDLSIMEEDKDGWT 153
 Db 62 GFTY--SDVTYTSQRKALAGVTFVALYDESRTELDLSFKGRRLQIVNTEGDWML 119
 QY 154 RVRRKEGEGGYPTSYL 170
 Db 120 AHSLTGTGTGYIPSNV 136
 RESULT 13
 SRC_AVI ST STANDARD; PRT; 557 AA.
 AC P14085;
 DT 01-JAN-1990 (Rel. 13 Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE TYROSINE-PROTEIN KINASE TRANSFORMING PROTEIN SRC (EC 2.7.1.112) (P60-
 DE SRC).
 GN V-SRC.
 OS Avian sarcoma virus (strain 52).
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
 OX NCBI_TaxId=11882;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87064539; PubMed=3097513;
 RA Ikawa S., Hagino-Yamagishi K., Kawai S., Yamamoto T., Toyoshima K.;
 RT "Activation of the cellular src gene by transducing retrovirus";
 RL Mol. Cell. Biol. 6:2420-2428(1986).
 CC -1- FUNCTION: THIS PHOSPHOPROTEIN, REQUIRED FOR BOTH THE INITIATION
 CC AND THE MAINTENANCE OF NEOPLASTIC TRANSFORMATION, IS A PROTEIN
 CC KINASE THAT CATALYZES THE PHOSPHORYLATION OF TYROSINE RESIDUES
 CC IN VITRO.
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE -> ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
 CC DOMAIN, BELONGS TO THE SRC SUBFAMILY.
 CC PIR: B25375; TYFVPS2.
 CC HSSP: P00523; 2PTK.
 CC InterPro: IPR000719; Euk_Pkinase.
 CC InterPro: IPR000980; SH2.
 CC InterPro: IPR001452; SH3.
 CC InterPro: IPR001245; Tyr_kin.
 CC Pfam: PF00069; pkinase; 1.
 CC Pfam: PF00017; SH2; 1.
 CC Pfam: PF00018; SH3; 1.
 CC PRINTS: PR00109; TYRKINASE.
 CC PRINTS: PR00401; SH2DOMAIN.
 CC PRINTS: PR00452; SH3DOMAIN.
 CC SMART: SM00252; SH2; 1.
 CC SMART: SM00326; SH3; 1.
 CC SMART: SM00219; TYRK; 1.
 CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 CC PROSITE: PS50001; SH2; 1.
 CC PROSITE: PS50002; SH3; 1.
 CC Tyrosine-protein kinase; Transforming protein; Oncogene;
 CC Transferrase; Phosphorylation; ATP-binding; Myristate;


```

DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS50001; SH2; 1.
DR PROSITE: PS50002; SH3; 1.
KW Tyrosine-protein kinase; Transforming protein; Oncogene;
KW Transferase; Phosphorylation; ATP-binding; Myristate;
KW SH3 domain; SH2 domain.
FT LIPID 2 MYRISTATE.
FT DOMAIN 81 142 SH3.
FT DOMAIN 148 245 SH2.
FT DOMAIN 267 520 PROTEIN KINASE.
FT NP_BIND 273 281 ATP (BY SIMILARITY).
FT BINDING 295 295 ATP (BY SIMILARITY).
FT ACT_SITE 386 386 BY SIMILARITY.
FT MOD_RES 416 416 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 568 AA; 63632 MW; 13777DB121F0996 CRC64;

Query Match          11.9%; Score 108; DB 1; Length 568;
Best Local Similarity 22.6%; Pred. No. 0.6;
Matches 31; Conservative 27; Mismatches 53; Indels 26; Gaps 4;

QY 58 SRHARPPXP---ASAPDDSS-----SNSAQDIKESSEPEPSSESDPTIYT 102
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 3 SSKSRKPDPQSQRRLSPEDSTHGHPASQTPNKTAAPIHRTPRSFGVATPEPLFG 62
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 103 EFDDEFEDEEPPS-----IGHCAVIHYHEGSECTISNAEEEDLSIMEEDKDGDWT 153
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 63 GFNT--SPTVTISPORAGALACGVTTFFVALDYESTRTETDLSPKKERLIQIVNNEGDMWL 120
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
QY 154 RVRRKEGEGEVPTSYL 170
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 121 AHSLTGTGTGYLPNNYV 137
   | : : | : : | : : | : : | : : | : : | : : | : : | : : | : : |

RESULT 15
SRC_AVIS2
ID SRC_AVIS2 STANDARD; PRT; 587 AA.
AC P15054;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TYROSINE-PROTEIN KINASE TRANSFORMING PROTEIN SRC (EC 2.7.1.112) (P60-SRC).
GN V-SRC.
OS Avian sarcoma virus (strain PR2257).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
CX NCBI_TaxID=11879;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8904972; Pubmed=2463376;
RA Geryk J., Dezelee P., Barnier J.V., Svoboda J., Nehyba J.,
RA Karakoz I., Rynditch A.V., Yatsula B.A., Calothy G.;
RA "Transduction of the cellular src gene and 3' adjacent sequences in
RA avian sarcoma virus PR2257.";
RL J. Virol. 63:481-492(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX Yatsula B.A., Geryk J., Svoboda J., Rynditch A.V., Calothy G.,
RA Dezelee P.;
RL Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 146-249.
RA Holland D.R., Lunney E.A., Plummer M.S., Mueller W.T., McConnell P.,
RA Pavlovsky A., Para K.S., Shahrilpour A., Humblet C., Sawyer T.K.,
RA Rubin J.R.;
RL Submitted (MAY-1997) to the PDB data bank.
CC -!- FUNCTION: THIS PHOSHOPROTEIN, REQUIRED FOR BOTH THE INITIATION
CC AND THE MAINTENANCE OF NEOPLASTIC TRANSFORMATION, IS A PROTEIN
CC KINASE THAT CATALYZES THE PHOSPHORYLATION OF TYROSINE RESIDUES
CC IN VITRO.
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +

```

```
CC      PROTEIN TYROSINE PHOSPHATE.
CC      -!- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
CC      -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC      -!- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC
CC      DOMAIN. BELONGS TO THE SRC SUBFAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
CC      EMBL: M21526; AAA42583.1; -.
CC      EMBL: X51863; CAA36156.1; -.
CC      PIR: A30174; TVFVPR.
CC      PDB: 1BKL; 23-JUL-97.
CC      PDB: 1BKM; 07-JUL-97.
CC      InterPro: IPR000719; Euk_pkinase.
CC      InterPro: IPR000980; SH2.
CC      InterPro: IPR001452; SH3.
CC      InterPro: IPR001245; Tyr_kin.
CC      Pfam: PF00069; pkinase; 1.
CC      Pfam: PF00017; SH2; 1.
CC      Pfam: PF00018; SH3; 1.
CC      PRINTS: PR00109; TYRKINASE.
CC      PRINTS: PR00401; SH2DOMAIN.
CC      PRINTS: PR00452; SH3DOMAIN.
CC      SMART: SM00252; SH2; 1.
CC      SMART: SM00326; SH3; 1.
CC      SMART: SM00319; Tyrc; 1.
CC      PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC      PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
CC      PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC      PROSITE: PS50001; SH2; 1.
CC      PROSITE: PS50002; SH3; 1.
CC      Tyrosine-protein kinase; Transforming protein; Oncogene;
CC      transferase; Phosphorylation; ATP-binding; Myristate;
CC      SH3 domain; SH2 domain; 3D-structure.
CC      LIPID 2 2 MYRISTATE.
CC      FT DOMAIN 81 142 SH3.
CC      FT DOMAIN 148 245 SH2.
CC      FT DOMAIN 267 520 PROTEIN KINASE.
CC      NP_BIND 273 281 ATP (BY SIMILARITY).
CC      FT BINDING 295 295 ATP (BY SIMILARITY).
CC      ACT_SITE 386 386 BY SIMILARITY.
CC      MOD_RES 416 416 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC      SEQUENCE 587 AA; 65800 MW; 0A6925315EF251D9 CRC64;

Query Match 11.9%; Score 108; DB 1; Length 587;
Best Local Similarity 22.6%; Pred. No. 0.62;
Matches 31; Conservative 27; Mismatches 53; Indels 26; Gaps 4;

OY 58 SRHARPPXP-----ASAPDSS-----SNSASQDTKESSEPPSESDTPPIYT 102
DB 3 SAKSKPKDPSQRRSLPEPDSTHGGFPASQTPNKTAAPDTHRTPTSPRSFGTVATEPKLFG 62
OY 103 EFDEDEFEETPTSP-----IGHCVAIYHFGSSSEGTISMAEGDLSIMEEDKGDGWT 153
DB 63 GENT--SDVTPTSPQAGALAGVTTFVALDYESTRTDLSFKKGRLOIVNNTGDMWL 120
OY 154 RVRKKEGEGYVPTSYL 170
DB 121 AHSLTGTGTGYIPSNV 137
```

Search completed: April 7, 2002, 16:14:12
Job time: 325 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 7, 2002, 16:13:11 ; Search time 40.07 Seconds

(without alignments)
638.823 Million cell updates/sec

Title: US-09-925-122a-3

Sequence: 1 MKDYKPTQMGDPASLEPQ.....RRKGEGGYPTSLRVTLN 175

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_invertebrate:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp_virus:*
13: sp-vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	908	99.8	545	4 O15184	O15184 homo sapien
2	812	89.2	547	11 P97531	P97531 rattus norv
3	804	88.4	547	11 O99L10	O99L10 mus musculu
4	450	49.5	330	4 O9NKG1	O9NKG1 homo sapien
5	450	49.5	434	4 O9BR31	O9BR31 homo sapien
6	432.5	47.5	537	4 O9NMD1	O9NMD1 homo sapien
7	432.5	47.5	592	4 O9H8H8	O9H8H8 homo sapien
8	432.5	47.5	674	4 O60301	O60301 mus musculu
9	422.5	46.4	237	11 O61053	O61053 mus musculu
10	302	33.2	525	5 O9VZD7	O9VZD7 drosophila
11	221	24.3	554	5 O19253	O19253 caenorhabdi
12	178	19.6	774	5 O9VSU8	O9VSU8 drosophila
13	177.5	19.5	783	5 O9U3B8	O9U3B8 caenorhabdi
14	169.5	18.6	785	5 O9XU57	O9XU57 caenorhabdi
15	156.5	17.2	603	4 O9UF77	O9UF77 homo sapien
16	156.5	17.2	684	4 O94868	O94868 homo sapien
17	131.5	14.5	701	4 O9NXX8	O9NXX8 homo sapien
18	128.5	14.1	448	13 O13154	O13154 gallus gall
19	123.5	13.6	335	4 O9Y4V2	O9Y4V2 homo sapien

20	123	13.5	445	4 O9H0D3	O9H0D3 homo sapien
21	123	13.5	445	11 O9GY17	O9GY17 rattus norv
22	121	13.3	447	11 O9GY19	O9GY19 rattus norv
23	120.5	13.2	486	11 O9WVE8	O9WVE8 mus musculu
24	119.5	13.1	687	11 O9GY53	O9GY53 mus musculu
25	119.5	13.1	691	11 O9GZP0	O9GZP0 mus musculu
26	118.5	13.0	486	11 O9GY18	O9GY18 rattus norv
27	118	13.0	731	4 O14837	O14837 homo sapien
28	118	13.0	732	4 O15259	O15259 homo sapien
29	116.5	12.8	488	4 O9UNF0	O9UNF0 homo sapien
30	116.5	12.8	488	11 O9GY20	O9GY20 rattus norv
31	115	12.6	368	3 O9P837	O9P837 candida alb
32	114.5	12.6	600	5 O9VE96	O9VE96 drosophila
33	114.5	12.6	739	11 O35613	O35613 mus musculu
34	113	12.4	477	13 O9DDA9	O9DDA9 xenopus lae
35	113	12.4	517	5 O77050	O77050 anthocidari
36	112	12.3	1217	3 O9Y7Z8	O9Y7Z8 schizosacch
37	110.5	12.1	502	13 O9DDK6	O9DDK6 salmo salar
38	110	12.1	111	4 O9NPN1	O9NPN1 homo sapien
39	110	12.1	740	11 O9QWT8	O9QWT8 mus musculu
40	109.5	12.0	1119	3 O9P3N5	O9P3N5 neurospora
41	109	12.0	424	4 O9UKS6	O9UKS6 homo sapien
42	109	12.0	424	4 O9H331	O9H331 homo sapien
43	109	12.0	424	11 O9EOP9	O9EOP9 mus musculu
44	109	12.0	686	11 O9WUZ2	O9WUZ2 mus musculu
45	108.5	11.9	498	10 O9ZOK6	O9ZOK6 arabidopsis

ALIGNMENTS

RESULT 1
O15184 PRELIMINARY; PRT; 545 AA.

AC O15184;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CDC42-INTERACTING PROTEIN 4.
GN CIP4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:97362357; Pubmed:9210375;
RA Asperstroem P.;
RT "A Cdc42 target protein with homology to the non-kinase domain of FER
RT has a potential role in regulating the actin cytoskeleton.";
RL Curr. Biol. 7:479-487(1997).
DR EMBL: AJ000414; CA004062.1; -
DR InterPro: IPR001060; FCH.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00018; SH3; 1.
DR Pfam: PF00611; FCH; 1.
DR PROSITE: PSS0002; SH3; 1.
DR SMART: SM00055; FCH; 1.
DR SMART: SM00326; SH3; 1.
SQ SEQUENCE 545 AA; 62591 MW; 9C9D72EA734BC62 CRC64;

Query Match 99.8%; Score 908; DB 4; Length 545;
Best Local Similarity 99.4%; Pred. No. 8.3e-72;
Matches 174; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MKDYKPTQMGDPASLEPQIAETLSNTERKLEVQKYEAFLAEASRYLSNRGDSLSRH 60
|||||
DB 371 MKDYKPTQMGDPASLEPQIAETLSNTERKLEVQKYEAFLAEASRYLSNRGDSLSRH 430
OY 61 ARPPPPASAPPPSSSSASODTKESSEPPSEESODTPITTEPDEDEEPTSPICHCY 120
|||||
DB 431 ARPPPPASAPPPSSSSASODTKESSEPPSEESODTPITTEPDEDEEPTSPICHCY 490

QY 121 ATYHFGSSEGTISMAEGEDLSLMEEDKDGWTRVRKKEGEGYVPTSYLRTLN 175
 DB 491 ATYHFGSSEGTISMAEGEDLSLMEEDKDGWTRVRKKEGEGYVPTSYLRTLN 545

RESULT 2

P97531 PRELIMINARY: PRT: 547 AA.

ID P97531
 AC P97531
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE SALT-TOLERANT PROTEIN.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxId=10116;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=WISTAR; TISSUE=KIDNEY;
 RA Tsuji E.;
 RL Submitted (Aug-1997) to the EMBL/Genbank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WISTAR; TISSUE=KIDNEY;
 RX MEDLINE=97112415; PubMed=8954095;
 RA Tsuji E., Tsuji Y., Misumi Y., Fujita A., Sasaguri M., Ideishi M.,
 RA Arkawa K.;
 RT "Molecular cloning of a novel rat salt-tolerant protein by functional
 RT complementation in yeast."
 RL Blochem. Biophys. Res. Commun. 229:134-138(1996).
 DR EMBL: AB006914; BAA22191.1; -
 DR InterPro: IPR001060; FCH.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00018; SH3; 1.
 DR Pfam: PF00611; FCH; 1.
 DR PROSITE: PS50002; SH3; 1.
 DR SMART: SM00055; FCH; 1.
 DR SMART: SM00326; SH3; 1.
 SQ SEQUENCE 547 AA; 62787 MW; 16CF72FDCAA66B3F CRC64;

Query Match 89.2%; Score 812; DB 11; Length 547;
 Best Local Similarity 87.6%; Pred. No. 2.3e-63;
 Matches 156; Conservative 7; Mismatches 11; Indels 4; Gaps 2;

QY 1 MKDYVETKTPOMGDPASLEPQIAETLSNIRKLEVKYKAMLAESA RYLSNRGDSLSRH 60
 DB 371 MKDYVETKTPOMGDPASLEPQIAETLSNIRKLEVKYKAMLAESA RYLSNRGDSLSRH 430
 QY 61 ARPPXPASAPPDSSS---NSASODTKESSEPPSEESODPTIYTEFDEDFEEDPTSPIG 117
 DB 431 TRPPPTTAPPDSSSSNSGSDNMESSEPPSEEGODPTIYTEFDEDFEEDPTSPIG 489
 QY 118 HCAVATYHFGSSEGTISMAEGEDLSLMEEDKDGWTRVRKKEGEGYVPTSYLRTLN 175
 DB 490 QCVATYHFGSSEGTISMAEGEDLSLMEEDKDGWTRVRKKEGEGYVPTSYLRTLN 547

RESULT 3

Q99L10 PRELIMINARY: PRT: 547 AA.

ID Q99L10
 AC Q99L10
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE STIMULIN TO THYROID HORMONE RECEPTOR INTERACTOR 10.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=40090;
 RN (1)

RP SEQUENCE FROM N.A.

RL Strausberg R.;

DR EMBL: BC003249; AAH03249.1; -

QY SEQUENCE 547 AA; 62653 MW; 5DE65DB3D6C9B8E2 CRC64;

Query Match 88.4%; Score 804; DB 11; Length 547;
 Best Local Similarity 87.6%; Pred. No. 1.1e-62;
 Matches 156; Conservative 5; Mismatches 13; Indels 4; Gaps 2;

QY 1 MKDYVETKTPOMGDPASLEPQIAETLSNIRKLEVKYKAMLAESA RYLSNRGDSLSRH 60
 DB 371 MKDYVETKTPOMGDPASLEPQIAETLSNIRKLEVKYKAMLAESA RYLSNRGDSLSRH 430
 QY 61 ARPPXPASAPPDSSS---NSASODTKESSEPPSEESODPTIYTEFDEDFEEDPTSPIG 117
 DB 431 TRPPPTTAPPDSSSSNSGSDNMESSEPPSEEGODPTIYTEFDEDFEEDPTSPIG 489
 QY 118 HCAVATYHFGSSEGTISMAEGEDLSLMEEDKDGWTRVRKKEGEGYVPTSYLRTLN 175
 DB 490 QCVATYHFGSSEGTISMAEGEDLSLMEEDKDGWTRVRKKEGEGYVPTSYLRTLN 547

RESULT 4

Q9NXG1 PRELIMINARY: PRT: 330 AA.

ID Q9NXG1
 AC Q9NXG1
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CDNA FLJ20275 FIS, CLONE HEP02372.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxId=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
 RA Nakajima Y., Mizuno T., Morinaga M., Ota T., Suzuki Y., Obayashi M.,
 RA Nishi T., Shibahara T., Tanaka T., Nakamura Y., Isegawa T., Sugano S.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AK000282; BAA91051.1; -
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS50002; SH3; 1.
 SQ SEQUENCE 330 AA; 37427 MW; 2F6260C9FD3C46AF CRC64;

Query Match 49.5%; Score 450; DB 4; Length 330;
 Best Local Similarity 50.3%; Pred. No. 7.9e-32;
 Matches 88; Conservative 30; Mismatches 49; Indels 8; Gaps 4;

QY 1 MKDYVETKTPOMGDPASLEPQIAETLSNIRKLEVKYKAMLAESA RYLSNRGDSLSRH 60
 DB 156 MKDYVETKTPOMGDPASLEPQIAETLSNIRKLEVKYKAMLAESA RYLSNRGDSLSRH 212
 QY 61 ARPPXPASAPPDSSSNSASODTKESSEPPSEESODPTIYTEFDEDFEEDPTSPIGHC 119
 DB 213 SSDIHLVYQGREPSGSDTDANQEVRRPQOHG---HNEFDEDFEEDPTSPIGHC 268
 QY 120 VAIYHFGSSEGTISMAEGEDLSLMEEDKDGWTRVRKKEGEGYVPTSYLRTLN 174
 DB 269 KAIYFEDGHNEGTLMKKEGEVLIYEEDKDGWTRVRKKEGEGYVPTSYLRTLN 323

RESULT 5

Q9BR51 PRELIMINARY: PRT: 434 AA.

ID Q9BR51
 AC Q9BR51;

```

DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE DJ103H22.1 (KIAA0554 PROTEIN) (FRAGMENT).
GN DJ103H22.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wallis J.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL109613; CAC36351.1; -.
FT NON_TER
SQ SEQUENCE 434 AA; 49389 MW; 2B12549604980040 CRC64;

Query Match
Best Local Similarity 49.5%; Score 450; DB 4; Length 434;
Matches 88; Conservative 30; Mismatches 49; Indels 8; Gaps 4;

QY 1 MKDYEKTPQMGDPASLEPOIAETLSNTERLKLEVOKYEAMLAESRVLNRCDSLSRH 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 256 MKDYEKTPQMGDPASLEPOIAETLSNTERLKLEVOKYEAMLAESRVLNRCDSLSRH 312
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 ARPPXPPASAPPPSSNSASODTRKESSEPPESODTPRYTEDEDFE-EEPTSGHC 119
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 313 SSDINHLVTOGRESPESSYTDANQEVGRPPQHG-NEEDDEDDDDPLPAIGHC 368
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 120 VAHYFEGSGEITSMAGEDLSMEEDKGDGWTVRKRKEGEGVPTSYLRTL 174
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 369 KAIFPGHNEGTLAMKEGELYIIEEDKGDGWTVRKRKEGEGVPTSYLRTL 423
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
Q9NMD1 PRELIMINARY; PRT; 537 AA.
ID 09NMD1:
AC 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE HYPOTHETICAL 61.6 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-EMBRYO;
RC Isogal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Watanabe M., Hosokawa T., Kaku Y., Kodaira H., Kondo Y., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Niinomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK000975; BAB91451.1; -.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00018; SH3; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PSS0002; SH3; 1.
SQ SEQUENCE 537 AA; 61560 MW; CE14592678DD1A65 CRC64;

Query Match
Best Local Similarity 47.5%; Score 432.5; DB 4; Length 537;
Matches 91; Conservative 28; Mismatches 46; Indels 19; Gaps 4;

QY 1 MKDYEKTPQMGDPASLEPOIAETLSNTERLKLEVOKYEAMLAESRVLNRCDSLSRH 55
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

DB 354 MKDYELKNPQMGDPASLDHKLAEVSONIEKLKLVETQFEAMLAEGRLPARSEQARRQS 413
QY 56 SLSHARPPAPPPASAPPPSSNSASODTRKESSEPPESODTPRYTEDEDF-EE 111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 414 GLVDSQNP-----TVNCAQDRESPDGSYTEBOSQSEKMLAFDFDEDFDE 463
QY 112 PTPSPGHCAIYHFGSGSGTISMAEGEDLSMEEDKGDGWTVRKRKEGEGVPTSYLR 171
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 464 PLPAIGTCALYTFEGNEGTSIVVEGETLYIIEEDKGDGWTVRKRNEDEGVPYSYE 523
QY 172 VTLN 175
    |||||
DB 524 VCLD 527

RESULT 7
Q9H8H8 PRELIMINARY; PRT; 592 AA.
ID 09H8H8:
AC 09H8H8;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE CDNA FLJ13619 FTS, CLONE PLACEL010926, WEAKLY SIMILAR TO HYPOTHETICAL
DE 72.2 KD PROTEIN C12C2.05C IN CHROMOSOME 11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-PLACENTA;
RC Isogal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K.,
RA Tanai H., Kimata M., Watanabe M., Hiraoaka S., Ishii S., Kawai Y.,
RA Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K.,
RA Masuho Y., Kanehori K.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK023681; BAB14638.1; -.
DR InterPro: IPR001060; FCH.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00611; FCH; 1.
DR SMART: SM00055; FCH; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PSS0002; SH3; 1.
SQ SEQUENCE 592 AA; 68910 MW; 0CAC9D464706B67 CRC64;

Query Match
Best Local Similarity 47.5%; Score 432.5; DB 4; Length 592;
Matches 91; Conservative 28; Mismatches 46; Indels 19; Gaps 4;

QY 1 MKDYEKTPQMGDPASLEPOIAETLSNTERLKLEVOKYEAMLAESRVLNRCDSLSRH 55
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 409 MKDYELKNPQMGDPASLDHKLAEVSONIEKLKLVETQFEAMLAEGRLPARNEQARRQS 468
QY 56 SLSHARPPAPPPASAPPPSSNSASODTRKESSEPPESODTPRYTEDEDF-EE 111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 469 GLVDSQNP-----TVNCAQDRESPDGSYTEBOSQSEKMLAFDFDEDFDE 518
QY 112 PTPSPGHCAIYHFGSGSGTISMAEGEDLSMEEDKGDGWTVRKRKEGEGVPTSYLR 171
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 519 PLPAIGTCALYTFEGNEGTSIVVEGETLYIIEEDKGDGWTVRKRNEDEGVPYSYE 578
QY 172 VTLN 175
    |||||
DB 579 VCLD 582

RESULT 8
O60301 PRELIMINARY; PRT; 674 AA.
ID O60301:

```

AC 060301;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE KIA00554 PROTEIN (FRAGMENT).
 GN KIA00554.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE=98290545; PubMed=9628581;
 RA Nagase T., Ishikawa K., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.";
 RL DNA Res. 5:31-39(1998).
 DR EMBL: AB011126; BAA25480.1; .
 DR InterPro: IPR001060; FCH.
 DR Pfam: PF00018; SH3; 1.
 DR Pfam: PF00611; FCH; 1.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR PROSITE: PS00002; SH3; 1.
 DR SMART: SM00055; FCH; 1.
 DR SMART: SM00326; SH3; 1.
 FT NON_TER
 FT SEQUENCE 674 AA; 77447 MW; DOB53BE0B2B8DE2F CRC64;

Query Match 47.5%; Score 432.5; DB 4; Length 674;
 Best Local Similarity 49.5%; Pred. No. 6e-30;
 Matches 91; Conservative 28; Mismatches 46; Indels 19; Gaps 4;
 QY 1 MKDYEKTPQMDPASPDLQIAETLSNIERKLEVKYKFAWLAESRY-----LSNRGD 55
 DB 495 MKDYLNKPMQMDPASPDLKLAESQNIETKLVETQKFAWLAESGRLPARSEQARRQS 554
 QY 56 SLRHARPPXPAPSPDSSNSASQDTKE--SSEPPSESDPTPTYTEDEDF-EET 111
 DB 555 GLYDQNP-----TWNCAQDRSPDGYTTEQSQSEKAVLATDDEDEDE 604
 QY 112 PTSPIGHCAVYHFEESSEGTISMAEGEDLSLMEEDKGGDGTVRKKEGEGVPTSYLR 171
 DB 605 PLPAIGTCKALYTFEGQNEGTVSEGETLVYIEEDKGGDGTVRIRNDEDEEGVPTSYVE 664
 QY 172 VTLN 175
 DB 665 VCLD 668
 RESULT 9
 ID 061053 PRELIMINARY; PRT; 237 AA.
 AC 061053;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE FORMIN BINDING PROTEIN FBP 17 (FRAGMENT).
 GN FMBP1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB;
 RX MEDLINE=96183189; PubMed=8605874;
 RA Chan D.C., Redford M.T., Leder P.;
 RT "Formin binding proteins bear WW/PW domains that bind proline-rich

RT peptides and functionally resemble SH3 domains.";
 RL EMBL: J. 15:1045-1054(1996).
 DR EMBL: U40751; AAC52479.1; .
 DR HSSP: P15054; 1BK.
 DR MGD: MGI:109606; Fmbp1.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00018; SH3; 1.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS00002; SH3; 1.
 FT NON_TER
 FT SEQUENCE 237 AA; 27173 MW; 4E23794C8CC744D7 CRC64;
 Query Match 46.4%; Score 422.5; DB 11; Length 237;
 Best Local Similarity 49.2%; Pred. No. 1.4e-29;
 Matches 88; Conservative 27; Mismatches 55; Indels 9; Gaps 3;
 QY 1 MKDYEKTPQMDPASPDLQIAETLSNIERKLEVKYKFAWLAESRYLSNRGDSLSRH 60
 DB 54 MKDYLNKPMQMDPASPDLQKLTVTQNIETKRLKLAQKFAWLAESGRLPARSEQ 108
 QY 61 ARPPXPAPSPDSSNSASQDTKE--SSEPPSESDPTPTYTEDEDF-EETPTSP 116
 DB 109 ARROSLGYDQTHQVTNCAQDRSPDGYTTEQSQSEKAVLATDDEDEDEPLPAI 168
 QY 117 GHCAVYHFEESSEGTISMAEGEDLSLMEEDKGGDGTVRKKEGEGVPTSYLRVTLN 175
 DB 169 GTCALYTFEGQNEGTVSEGETLVYIEEDKGGDGTVRIRNDEDEEGVPTSYVEYLD 227
 RESULT 10
 ID 09VZD7 PRELIMINARY; PRT; 525 AA.
 AC 09VZD7;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE CG15015 PROTEIN.
 GN CG15015.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkov D., Botchan M.R., Bouck J., Brokstein P., Brotlier P., Buttis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W., Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jaimel M.E., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Lin X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copest T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier J., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightnig J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterson R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.; III of C.
RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*.";

RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller K.M., Basu A., Baxendale J., Bayraktaroglu I., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ileguam C.,
 RA Jalali M., Kaulsh F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Mostneff A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Massarman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*";
 RA Science 287:2185-2195(2000).
 RA EMBL: AE003553; AAF5013.1; -.
 RA HSSP: P29354; 1GRT.
 DR Flybase: FBgn003596; CG4684.
 DR InterPro: IPR001060; FCH.
 DR InterPro: IPR00108; Neu_cyl_fact_2.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00018; SH3; 2.
 DR PRINTS: PR00499; P67PHOX.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR SMART: SM00055; FCH; 1.
 DR SMART: SM00326; SH3; 2.
 DR PROSITE: PS50002; SH3; 2.
 DR NON_TER 774
 FT 774
 SQ SEQUENCE 774 AA; 87266 MW; 69641BICE27FBCC CRC64;

Query Match 19.6%; Score 178; DB 5; Length 774;
 Best Local Similarity 22.9%; Pred. No. 1.6e-07;
 Matches 55; Conservative 36; Mismatches 81; Indels 68; Gaps 6;
 Oy 1 MKDYERT-PQMGDPASLEPOIAETLSNIERKLEVOYKAE-----WLAEA 45
 Db 361 LRDSGRTDPNDPNCPLDPTKIEFRDOIRSETEKTKAEACLOCLRDGGINDEVWQEA 420
 Oy 46 ES---RVLSNRGDSLSRHRAPRXPASAPPPSSNSASODTKSESPPESESO----- 96
 Db 421 ENMGVELTRSSASISMRIDASGGENPSSDSRYDSDEKTOAAOQKPKQEOQLSRDRT 460
 Oy 97 -----DTPIYTEP-----DEDFEEPTSP-- 115
 Db 481 FSDSEDEPEVRPSAAASAAASAAASMMASAGGMDPTEVNMAGAGEEDKDEPIVPEP 540
 Oy 116 ---IGHCVAIYFEGSGSEGTISMAEGEDLSLMEEDKGDGWTVRKRKEGGGCVPTSYLR 172
 Db 541 KEAIFCTALYSTAQNPELITVENOLEVVEGGDGLRARNYRGEGYVPHNYLDI 600

RESULT 13
 ID 0903B8 PRELIMINARY; PRT; 783 AA.
 AC 0903B8;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE K08E3.3A PROTEIN.

GN K08E3.3A.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderiinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN 11
 RP SEQUENCE FROM N.A.
 RA McMuray A.A.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN 12
 RP SEQUENCE FROM N.A.
 RA MEDLINE=99069613; PubMed=9851916;
 RX none;
 RA none;
 RT "genome sequence of the nematode *C. elegans*: A platform for
 RT investigating biology";
 CC -1- SIMILARITY: TO EF-HAND FAMILY.
 RL Science 282:2012-2018(1998).
 DR EMBL: Z81568; CAB04591.1; -.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR001060; FCH.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00036; efhand; 1.
 DR Pfam: PF00611; FCH; 1.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00055; FCH; 1.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS50002; SH3; 1.
 KW Calcium-binding.
 SQ SEQUENCE 783 AA; 88330 MW; 466AECA3E5FCA2 CRC64;

Query Match 19.5%; Score 177.5; DB 5; Length 783;
 Best Local Similarity 24.8%; Pred. No. 1.7e-07;
 Matches 55; Conservative 32; Mismatches 82; Indels 53; Gaps 5;
 Oy 1 MKDYERT-PQMGDPASLEPOIAETLSNIERKLEVOYKAEWLAESRSLNRGDSLSRH 60
 Db 561 LQOAYYTPNPGHNSACTEPLISYAKKIEKIMHMKEFYALWESV--EEGQERSFG 618
 Oy 61 ARPPXPPASAPPPSSNSASODTKES-----SEE 89
 Db 619 GRTPTTRSMGSGSTINQSSKTIEDVLSGEAGNSSADDSKNILROLFTTPKRLISSP 678
 Oy 90 PPSEESQDTP-----IYTFEDEFEEPTSPIGHCV-----AIYFEGSGSE 130
 Db 679 KTKSSPTPLRRRAETLSSPKILRSSGAIKRSLSITPDSVKVETAVYALPEFASSA 738
 Oy 131 GTISMAEGEDLSLMEEDKGDGWTVRKRKEGGE-GYVPTSYLR 171
 Db 739 ETWSIEGELIVLEHHDGDMGTTRKKHNEESGCVPTSYLQ 780

RESULT 14
 ID 09XUS7 PRELIMINARY; PRT; 785 AA.
 AC 09XUS7;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE K08E3.3B PROTEIN.
 GN K08E3.3B.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderiinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN 11
 RP SEQUENCE FROM N.A.
 RA McMuray A.A.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN 12
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Koopa A., Saunders D., Showkhen R.,
 RA Smaildon N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
 RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkison-Sproat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 366:32-38(1994).
 CC - SIMILARITY: TO EF-HAND FAMILY.
 DR EMBL: Z81568; CAB04595.1; -;
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR001060; FCH.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00036; eHand; 1.
 DR Pfam: PF00611; FCH; 1.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00055; FCH; 1.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS00002; SH3; 1.
 KW Calcium-binding.
 SQ SEQUENCE 785 AA; 88576 MW; FDDC36C4EA728508 CRC64;

Query Match 18.6%; Score 169.5; DB 5; Length 785;
 Best Local Similarity 25.0%; Pred. No. 8.8e-07;
 Matches 56; Conservative 31; Mismatches 82; Indels 55; Gaps 6;

QY 1 MKDVEKTPQMDPASPASLEPOIAETLSNIEKLEVKYEAFLAEESRYLSNRGDSLRRH 60
 DB 561 LQAAVYTHGHPGNSPACTPCTSLYAKIKELKMDINKLKEFYAMLEMSV--EEGQERSFG 618
 QY 61 ARPPXPAPASAPDDSSNSASQDTKES-----SEE 89
 DB 619 GRDTPDPTSRMSGSSSTNOSSKTIEDVLSGEAGNSSADSSKNILROLFTTPKRLISSP 678
 QY 90 PSEESQDTP-----IYTFEDDFEPEEPTPIGHCY-----AIYHFESSSE 130
 DB 679 KTSKSTPTPLRRRAEISPKILRSSFGAIKRSISTPDSVKYETAIVTALFEFAKSSA 738
 QY 131 GTISMAEGEDLSLMEEDKGDGWTFRV--RKEGGE-GYVPTSYLR 171
 DB 739 ETMSIEQGEILLVLEHGDGWTFRTRKCRKHNEESGFVPTSYLQ 782

RESULT 15

Q9UF77 PRELIMINARY; PRT; 603 AA.
 AC Q9UF77;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOTHETICAL 68.1 KDA PROTEIN (FRAGMENT).
 GN DKEZP434L127.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid:9606;
 RN 11
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RA Koehler K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL133567; CAB63720.1; -;
 DR HSSP: P19174; ZHSP.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00018; SH3; 2.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR SMART: SM00326; SH3; 2.
 DR PROSITE: PS00002; SH3; 2.

KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 603 AA; 68107 MW; 1F8762061DCEFA677 CRC64;

Query Match 17.2%; Score 156.5; DB 4; Length 603;
 Best Local Similarity 25.8%; Pred. No. 9.1e-06;
 Matches 46; Conservative 27; Mismatches 68; Indels 37; Gaps 5;

QY 15 ASLEPQAEFLSNIEKLEVKYEA-----WLAESRYLSNRGDSLRRH 60
 DB 228 AELEQKIDEARENIRKAEIILIKAEARLDLLKQIGVSDTWLKSAMQVMELEN--ERW 285
 QY 61 ARPPXPAPASAPDDSSNSASQDTKESSEEPSEOPTPIYTEDEDFEPEEPTPIG--- 117
 DB 286 ARPPA-----VTSNGTILSLNADTFEREGEFEEDNM-----DYFDDSSSSPSGTLR 331
 QY 118 ---HCVAIYHFESSSEGTISMAEGEDLSLMEEDKGDGWTFRVRKKEGGGYVPTSYLR 171
 DB 332 NYP LTCVYYSYKASQPDDELTFIEHEVLEVTEDGDMEDWVKARKKVGQVYVEKYLQ 389

Search completed: April 7, 2002, 16:13:12
 Job time: 330 sec

